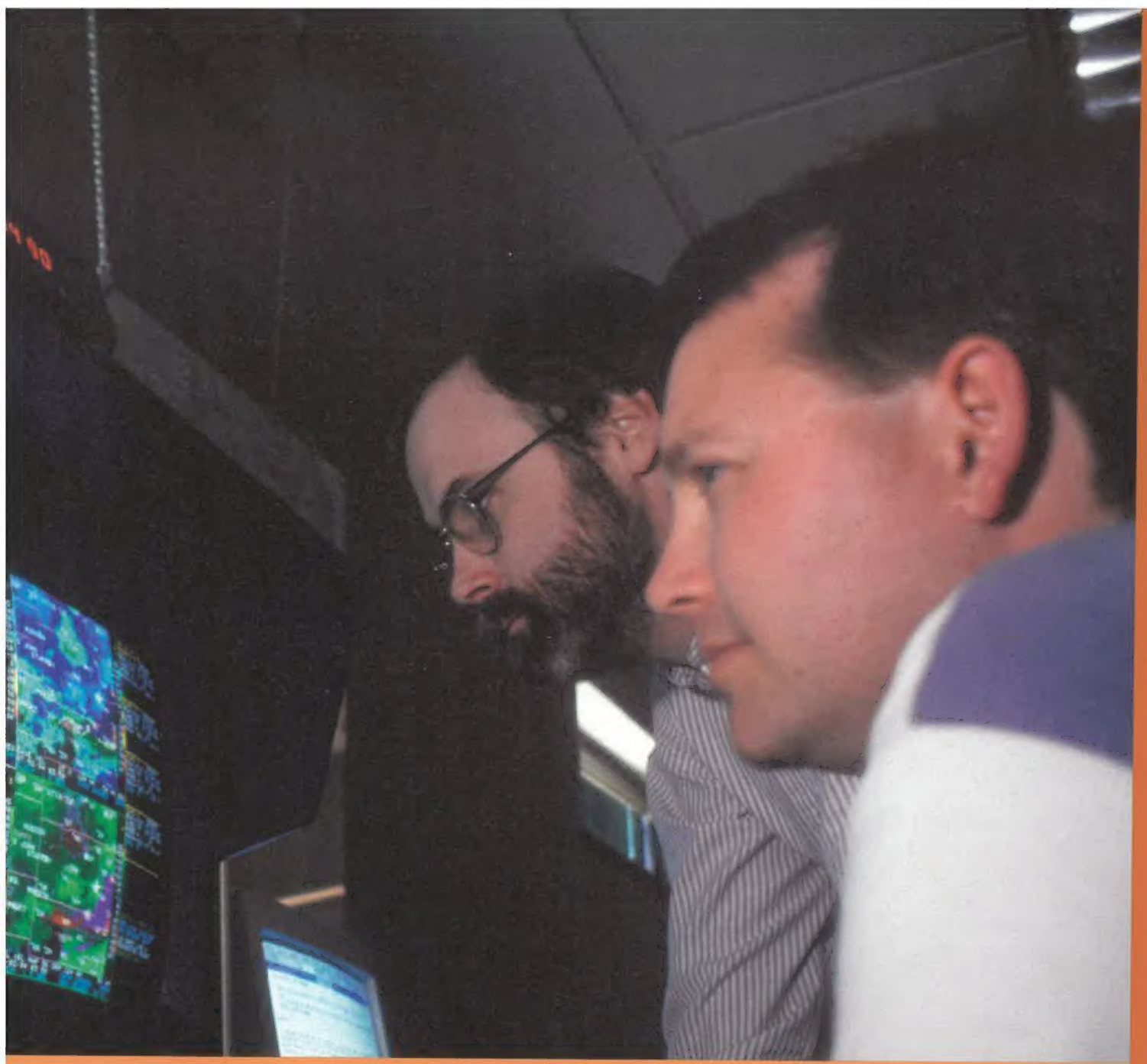


Data Modelling 5

Getting to know your unit

Assessment

You will be assessed by a series of assignments set by your tutor.

Every organisation is constantly faced with situations that require decisions, so producing alternative solutions to decision making is very important. Each decision has an impact, which can be positive if an effective decision is made, or negative if a poor decision is made. This unit explores how complex spreadsheets can support decision making in organisations by showing the consequences of alternative choices before they are actually made. You will find out how to design, develop and test a spreadsheet before responding to user feedback to make it better.

How you will be assessed

This unit will be assessed internally by means of a series of assignments set by your tutor. Throughout this unit, you will find assessment practice activities that will help you work towards your assignments. Completing these activities will not mean that you have achieved a particular grade, but you will have carried out useful research or preparation that will be relevant when it comes to your final assignment.

In order to complete the tasks in your assignment successfully, it is important to check that you have met all of the Pass grading criteria. You can do this as you work your way through the assignment.

If you are hoping to gain a Merit or Distinction, you should make sure that you present the information in your assignment in the style that is required by the relevant assessment criteria. For example, Merit criteria require you to analyse and discuss, and Distinction criteria require you to assess and evaluate.

The assignment set by your tutor will consist of a number of tasks designed to meet the criteria in the table below. The assignment is likely to consist of a written report but may also include:

- practical activities, to design and develop a data model to fulfil a client's requirements
- design documentation to plan your spreadsheet
- logs to record your spreadsheet development and testing.

Assessment criteria

This table shows what you must do in order to achieve a **Pass**, **Merit** or **Distinction** grade, and where you can find activities to help you.

Pass	Merit	Distinction
Learning aim A Investigate data modelling and how it can be used in the decision-making process		
A.P1 Explain the stages involved in the decision-making process for data modelling. **Assessment practice 4.1**	A.M1 Analyse how the features of spreadsheet software contribute to the decision-making process. **Assessment practice 4.1**	A.D1 Evaluate how the features of spreadsheet software contribute to the decision-making process. **Assessment practice 4.1**
A.P2 Explain how the features of spreadsheet software are used to support the decision-making process. **Assessment practice 4.1**		
Learning aim B Design a data model to meet client requirements		
B.P3 Produce designs for a data model which meets client requirements. **Assessment practice 4.2**	B.M2 Justify decisions made, showing how the design will fulfil its purpose and client requirements. **Assessment practice 4.2**	BC.D2 Evaluate the design and optimised data model against client requirements. **Assessment practice 4.2**
B.P4 Review the designs with others, to identify and inform improvements. **Assessment practice 4.2**		
Learning aim C Develop a data model to meet client requirements		
C.P5 Develop a data model to meet client requirements. **Assessment practice 4.2**	C.M3 Optimise the data model to meet client requirements. **Assessment practice 4.2**	BC.D3 Demonstrate individual responsibility, creativity, and effective self-management in the design, development and review of a data model. **Assessment practice 4.2**
C.P6 Test the data model for correctness, functionality and acceptance. **Assessment practice 4.2**		
C.P7 Review the extent to which the data model meets client requirements. **Assessment practice 4.2**		

Getting started

Data modelling involves using a spreadsheet or another app to mimic a real-life situation in which business decisions need to be made. It allows all options to be considered by predicting likely outcomes. Can you list some examples of real-life situations where a data model would be useful?

A Investigate data modelling and how it can be used in the decision-making process

Before you can design or create a spreadsheet to meet client requirements, you need to understand what data modelling is and how it is used. This section takes you through the stages in the decision-making process before moving on to the spreadsheet features that you will use to create your data models. Data modelling is used to help consider alternative choices. You will find out how to evaluate models and how to document your choices, justifying your decisions.

Stages in the decision-making process

There are many stages that you will explore in the decision-making process which include:

- understanding the scenario
- identifying information and sources
- recognising the factors affecting the quality of information
- analysing information
- identifying alternatives and the consequences of implementing them
- making and justifying decisions
- communicating decisions to others.

Understanding the scenario

The **scenario** is the situation in which the data model is needed. If you do not understand this there will inevitably be problems with the usefulness of your spreadsheet. In summary, the scenario lets you understand what your spreadsheet needs to be able to do.

Key term

Scenario – the situation or context within which you will need to create your data model.

The first stage in this process is to know who wants the spreadsheet, why they need it and what they want it to show. To understand the choices they need to make, your spreadsheet must be able to help their decision making by showing what is likely to happen from each possible alternative and therefore which of the choices is the best option. For example, if the decision-making process is being used to decide upon the price of a new product, the spreadsheet would need to be able to explore all the choices. These would include how many are made, how many are sold, quantities and costs of components, other costs (such as business premises or staff wages), the profit wanted for the product and the retail prices to the trade and the public.

Identifying information and sources

Every data model needs information, so, for this stage in the decision-making process, you will need to know where your information will come from and how reliable it will be.

Information required

This is the information required for the spreadsheet. You need to identify all the information needed for the spreadsheet data model including, for example, the profit to be made by producing a new product, or your model cannot work. When you know what information is needed, you can identify where it will come from.

Tip

You could list the information you have identified as being needed for your data model in one column of a spreadsheet and each of their sources in the next column. This can be used as a checklist to ensure that nothing is missed.

Information that is already available

Some of the information you need for your data model will already be available from the scenario. You should read the scenario carefully to identify this information.

For example, you may be able to identify the number of potential customers or the costs involved in producing the new product.

Additional information needed

There may also be additional information that your data model needs that is not given in the scenario. You will need to find out where this information can be found, that is, the source of this information. For example, if you are planning a data model that needs historical weather information for regions in the UK, then an internet search for 'download UK regional weather' would bring up the UK Met Office website, which provides a wide choice of downloadable data. In a product development scenario, the additional information needed could be how much are customers willing to pay for the product or the form of advertising that will be most effective.

Sources of additional information

You must identify the sources of any additional information that you will need for your spreadsheet data model. If these sources are from the internet, be careful to record exactly where they are located. There are likely to be lots of paper-based sources too. Try looking in the learning resource centre at your place of study or visit your local library. It might be possible to get the additional information that you need from primary sources such as interviews or questionnaires that you carry out yourself, but if you do, be very careful to make sure that you are querying the right people and that you have a good number of responses. For a new product, market research is often used to source additional information.

Discussion

The internet and paper-based sources are both good sources of information. What other sources can you identify? How reliable are these sources?

Requirements for verifying the information sources

All of the information you use for your spreadsheet needs to be trustworthy and reliable so that you can be confident that your data model will be useful and accurate. Verifying your information sources involves making sure that they are good and usable. To do this, find out who has provided the information and how old it is. If you have never heard of the information provider before, research them to find out who they are and why they have made the information accessible. Be especially cautious of organisations who are selling goods associated with the information. Also, most information gets less useful as it gets older, so always try to discover how old the data is. If it is two or more years old, think carefully about whether it is still relevant and good enough for your model. For a new product, verifying sources of information would involve identifying the core market for the product and gathering market research data from this group.

Factors affecting the quality of information

The currency and accuracy of the **data** are the most important factors affecting the quality of the **information**. It could also be influenced by other external factors.

Key terms

Data – the raw facts and figures, in an unorganised state. Often there is a massive amount of raw information and it is usually difficult to understand.

Information – produced when data is processed and presented so that it has meaning.

Currency of data

The currency of data is how recent it is. As stated above, older data tends to be less useful than up-to-date, recently produced data. Try to avoid old data that has lost its currency as it may make the information that you use for your data model a lot less useful.

Accuracy of data

The accuracy of data is how error free it is and also how detailed it is. Accurate data can be processed into a good and useful model. Try to avoid inaccurate data as the errors will produce poor results from your data model.

External factors

An external factor is something that is outside your data model but which may still influence it. External factors can be anything that may have an effect on how good your information is. For example, a car manufacturer could have a data model to predict the sales of their vehicles throughout the year based upon sales from the last decade. An external factor, such as a rival manufacturer bringing out a new model with very similar specifications at a much better price in the last two years, would make this data model very inaccurate.

Analysing the information

Analysing information is an important skill that is used by many IT professionals in a wide variety of jobs. These range from a helpdesk manager who needs to understand the support needs of the users of a network to a systems analyst planning a new software development project. So how is information analysed? The most important part of analysis is to understand what data needs to be collected and how it will be processed in order to find out the required information.

Worked example: Temperature and sunshine level changes

You are asked to produce a chart to identify trends in changes in temperature and sunshine through the years, on a monthly basis. The purpose of this data model is to produce information needed by an ice cream manufacturer to help them to decide how much ice cream to produce. What data is required to make this model work?

Step 1: Decide what each axis on the chart will be used for. For this model, use the *y*-axis for temperatures and sunshine hours and the *x*-axis for the months and years.

Step 2: Identify the data that is needed to make this work. You will need figures for both temperatures and sunshine hours for each month of each year. Ideally, this data should go back at least 50 years so will probably need to be be found in several separate data sets. The data sets need to be practical to import into the spreadsheet that you intend to use (eg Microsoft® Excel®).

Step 3: Identify where the data can be found. An internet search shows that the UK Met Office has some useful sets of weather data.

Step 4: Open the most suitable-looking data sets to check that they contains the required data.

Step 5: Download a data set to confirm that it can be used in Excel®.

Identifying alternatives

Always remember that any analysis of information that you produce is likely to be one of several possible solutions. Good practice is to carefully consider all the known alternatives before recommending or implementing a solution. To do this, you could try to approach the analysis from a different direction. Are there any other ways in which the information could be presented or calculated? Are there alternative data sources that you could use?

Identifying consequences of implementing the alternatives

Once you have identified alternative methods of information analysis and sources of data, you must be able to make a sensible decision about which to use.

A good starting point for making this choice is to identify the quality of the information produced from implementing each of the alternatives. What would be the consequences of using each of the alternatives? Would the information be any better? Is it more, or less, reliable or current? Would it take longer to produce the data model, or less? Is the data more, or less, expensive?

Making a decision

After identifying the consequences of implementing each of your alternatives, you should be able to understand which are the best and worst alternatives and therefore choose the one which has the best balance of quality and practicality.

Justifying the decision

You must be able to justify your choice of information. Understanding the consequences of implementing the alternatives should have provided good, solid reasons for why you made your choice. These reasons will need to be explained and justified as part of your assessment for this unit.

Communicating decisions

Communicating decisions to others is always important. It is essential for everyone involved in the data modelling project to be informed of your decisions, using the most appropriate methods of communication for each group of people.

Client

The client is the person from the scenario who wants to use your data model when it is ready. They know what is required and need to be kept up to date with your decisions. You may not communicate with the client every day but, instead, you may have regular meetings and give weekly or fortnightly reports. Good ways to communicate with the client are in person at meetings, in a report attached to an email or over the telephone. You may also consider sending letters if the client likes this form of communication.

Supervisor

Your supervisor needs regular updates on progress and an early heads-up or warning of any problems that develop. This is so that they can support you and are easily able to update anyone else who needs to know how the project is progressing. Most of the communication with your supervisor will probably be in the form of face-to-face conversations in the workplace. You may also communicate with them via email or telephone or at regular meetings.

Project sponsor

The project sponsor is the budget holder and is likely to be a manager in the organisation. Most communication with the project sponsor will be through your supervisor. Your supervisor, or you in certain circumstances, will communicate with the project sponsor in person at regular meetings, or through updates via email or telephone.

Spreadsheet features used to support data modelling

This section goes through the spreadsheet features that you will use to support your data modelling. The examples and techniques for this unit use the Microsoft® Excel® spreadsheet software app.

Entering and editing data

Data in a spreadsheet is kept in cells. Each cell looks like a rectangle in the spreadsheet. Every cell has a reference, so that it can be used in calculations. The cell reference is the column letter (from above the cell) and a number (from the left of the cell), for example B2.

Entering data is easy: simply select the cell then type. To complete the data entry you can tap the <Enter> key or use the <Arrow> keys to complete the data entry and move on to another cell in the direction of the arrow.

You edit data when you change something inside a cell. To do this, you need to select the cell then start the editing mode.

PAUSE POINT Can you explain the stages involved in the decision-making process for data modelling?

Hint Draw a diagram to show the sequence of the stages.

Extend What communications would be made at each stage? How would each be communicated and to whom would you send them?

You can also choose to replace what is in the cell by selecting the cell, then typing just as if the cell was empty. This is followed by pressing the <Enter> key or one of the <Arrow> keys.

Formatting data

There are many ways in which you can format the data in a spreadsheet, for example to change colour or appearance. First you need to select the cells containing the data. Then you can apply the formatting to the cells.

Link

See the Layout and presentation section for details on how to use formatting techniques.

Using formulae and functions

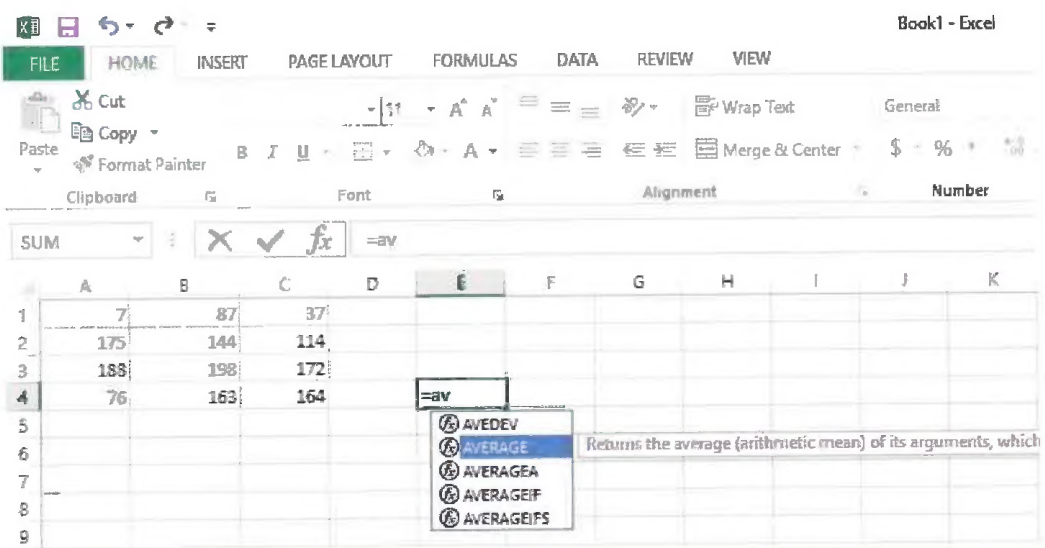
Formulae are calculations which start with = to let the spreadsheet know that the cell contains something that needs to be worked out. For example, typing = A2 * B4 would multiply the numbers in cells A2 and B4 together and the answer would be shown in the cell containing this formula.

Functions are more complex calculations which are already inside the spreadsheet, waiting to be used, for example =COUNTIF(B1:D1000,'Pass') which would show how many cells contain the text 'Pass' in the three thousand cells from B1 to D1000.

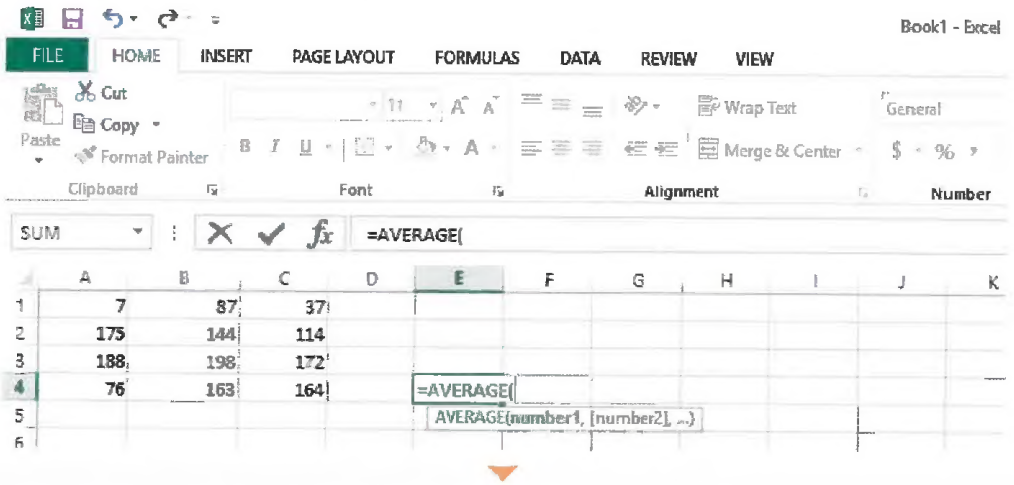
Step by step: Entering a function

4 Steps

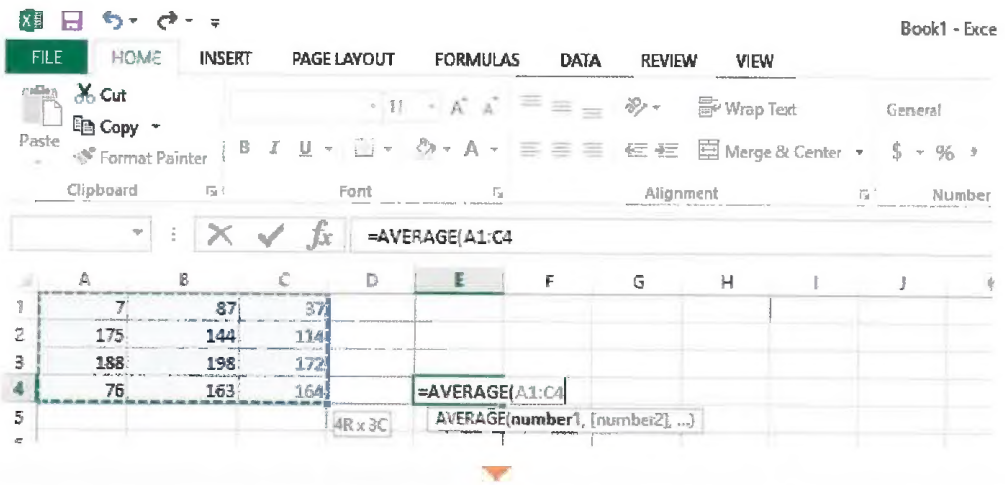
1 This step by step shows how to enter a function into cell E4 to find the average of the numbers in cells A1 to C4. Select the cell in which you want the function. Type = to start the function. Type the first few letters of the function. You will see a list of functions beginning with the letters you have typed. Use the <Down arrow> key to highlight AVERAGE.



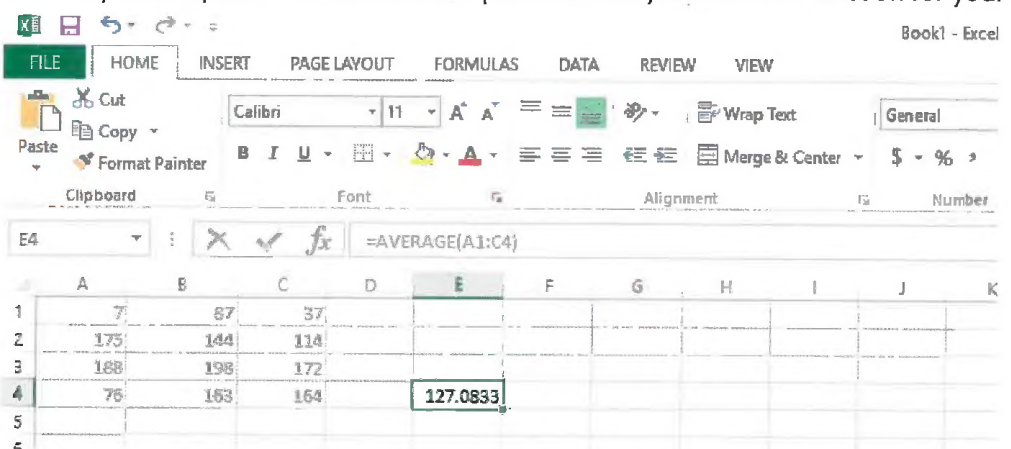
2 Use the <Tab> key to select AVERAGE. Notice how this also types the open bracket for you. Alternatively, you could double left click on AVERAGE at step 1.



3 Select the cells which are to be averaged. You can drag the mouse over these or use the <Arrow> keys to select one corner (C4) then hold a <Shift> key to highlight the required cells using the <Arrow> keys to move to the opposite corner (A1). Notice how there is a prompt under the function 'AVERAGE(number1,' (with what you are currently entering in bold), to let you know which cells the function will work on. In the example here this is A1:C4.



4 Use the <Enter> key to complete the task. Excel will put in the last) to finish the function for you.



Tip

You can jump around an Excel® spreadsheet quite easily using these keys:

- <Ctrl> <Home> to the top left corner
- <Home> to the left of the current row
- <Ctrl> <Arrow> to the edge of the current empty or filled cells in the direction of the <Arrow> key.

You can easily select parts of a spreadsheet by holding the <Shift> key while using any of the above.

Validation and verification of data

Many mistakes can be made when data is typed into a spreadsheet. Two ways of reducing data entry errors are validation and verification of data. The computer can help with validation and provides some support for verification.

Validation

Validation occurs when data entry is checked to make sure that it is OK (valid). Excel® allows you to set up validation for a cell to check for any of the following:

- whole numbers
- decimal numbers
- dates
- times
- text length
- data entries that are too big or too small (outside a range).

You can also use a list box to choose from pre-set choices for the data entry.

Excel® allows you to set up checks for data entry problems such as:

- duplicate entries
- spaces before, after or in the data entry.

There are many other ways in which you can set up calculations and check for data entry problems.

Link

See the Data entry and validation section on how to make these validation techniques work.

Verification

Verification is when data entry is checked to make sure that it is correct (verified). A computer does not have the intelligence to make sure that something typed in is actually correct. The best way to verify data entry using Excel® is to type the same data in twice. The second entry can then be checked against the first by an Excel® function or conditional formatting to warn the user if the two entries are different.

Analysing and interpreting data

You will need to carefully explore the data you find in order to get useful information from it. This is similar to analysing the information (as outlined in the section entitled Stages in the decision-making process), but here the analysis is from the bottom up. Previously, we started with the required information, analysed it to discover what data was needed and then processed it to produce the required information. Now you need to analyse the data to find out how to process or interpret it to get the required information.

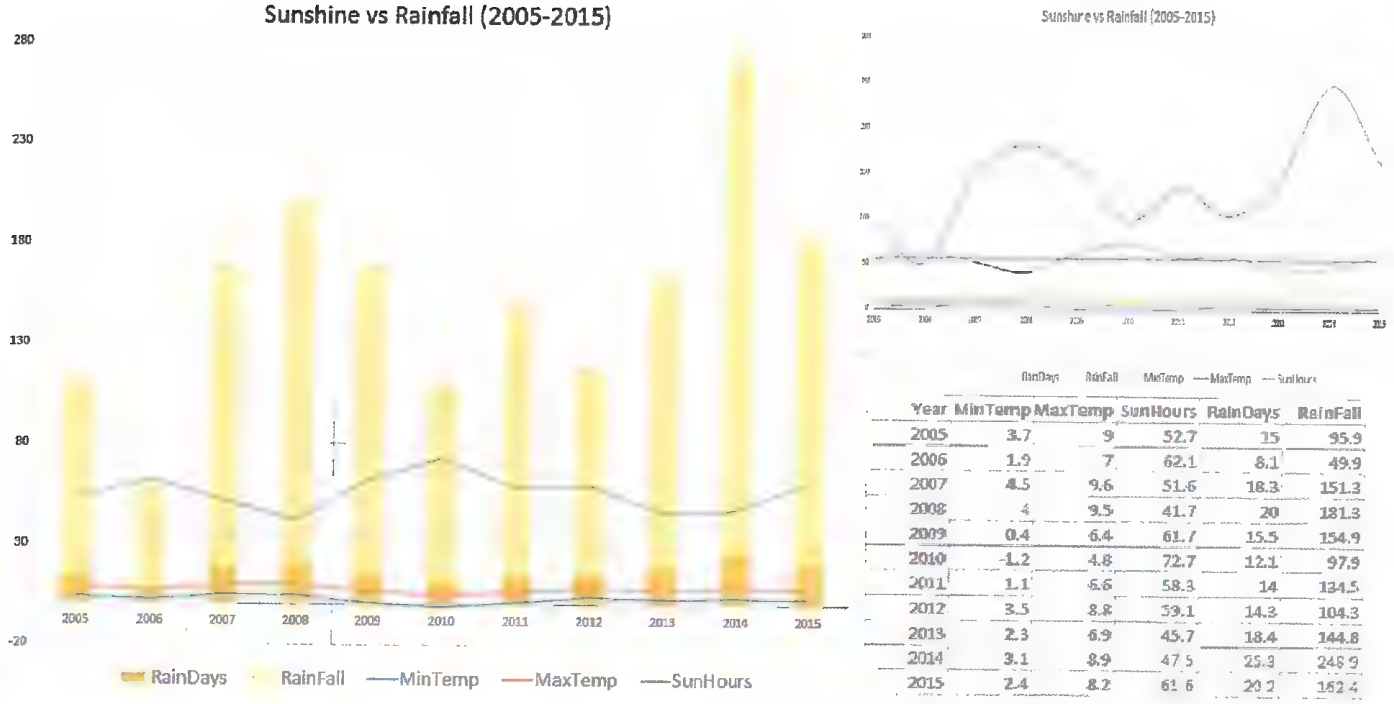
Data can be very large and often takes the form of an enormous table of numbers in a spreadsheet. An effective data model will analyse these numbers to interpret them into something more understandable. Sorting and filtering can bring some order and sense to large data sets. Functions can be used to count, find the average and carry out other processing actions, while conditional formatting can show interesting items and charts and can be a very effective way of helping people understand the meaning of the data. All of this requires some analysis or understanding of the data before the power of the spreadsheet can be used to bring it to life.

Presenting data

How data is presented is always important and a lot of time and skill can be required to get it right. There are often large quantities of data, for example, a hundred thousand items or more. Presentation is one of the tools used to bring meaning to this data.

As a general rule, less is more when presenting data. Simple summaries are often much more useful than lots of detail. A data modeller makes the calculations and processes the data to produce these summaries. It is a good idea to group related information, use font sizes and colour consistently, use charts where they add value and ensure that it is possible to print the summaries sensibly on a page.

Charts can be an excellent way to present data. Again, the skill here is to produce information which actually means something. Figure 5.1 shows how much easier it is to understand the larger combination chart in which the two sets of rain data are shown as stacked columns and the three sets of sunshine and temperature data are shown as lines. The smaller chart shows all five sets of data as lines and is not as easy to interpret.



Year	MinTemp	MaxTemp	SunHours	RainDays	RainFall
2005	3.7	9	52.7	15	95.9
2006	1.9	7	62.1	8.1	49.9
2007	4.5	9.6	51.6	18.3	151.3
2008	4	9.5	41.7	20	181.3
2009	0.4	6.4	61.7	15.5	154.9
2010	-1.2	4.8	72.7	12.1	97.9
2011	1.1	6.6	58.3	14	134.5
2012	3.5	8.8	59.1	14.3	104.3
2013	2.3	6.9	45.7	18.4	144.8
2014	3.1	8.9	47.5	25.3	248.9
2015	2.4	8.2	61.6	20.2	162.4

Figure 5.1: Same data, different charts (left: larger combination chart; right: smaller chart where all five data sets are lines)

PAUSE POINT

List the features of spreadsheet software and explain what they do.

Create a two-column table with features listed in one column and an explanation of what they do in the other.

Extend

Explain which spreadsheet functions can be used to help the decision-making process. How and why are they helpful?

Using data modelling to consider alternatives

When a data model has been set up, it can be used by the data modeller to identify possible alternatives and to choose the best of them. This is usually done by testing different input values.

Identifying the inputs required for the model

The inputs required for a model are used for the 'what if' analysis of possible alternatives, such as: What if we increase our prices by 20 per cent? The key here is to try out a range of possible inputs to find the balance needed to produce acceptable outcomes.

The range of outputs that can be produced

Outputs produced by the model could be viewed on-screen, printed or exported. On-screen is best for exploring possible input combinations, printed is best to share the best alternatives with others and exported is best for further reporting using other apps. A good data model will produce easily understandable outputs which clearly show the meaning of the data to meet the requirements of the client.

Alternative solutions

A good data model will be able to generate several possible solutions. Each of these will need careful consideration so as to justify your choice.

Benefits and limitations

What are the benefits and limitations of each alternative solution? These need to be carefully considered so as to identify and justify your preferred choice. You can produce a table showing the benefits and limitations of each alternative solution to help identify the most appropriate.

Impacts and consequences

What are the impacts or consequences of each alternative solution? Factors to consider are:

- risk (does a solution produce more or less)
- differences in timescales
- differences in costs.

Identifying the best decision or compromise

Your data model may not be able to find a perfect solution. You need to weigh up the benefits and limitations of the alternative solutions very carefully in order to find the one that fits the users' needs best or involves the least amount of compromise.

Evaluating models

You need to consider many factors (see Figure 5.2) when evaluating your model including:

- how well it reflects the scenario
- the decisions it supports
- whether it can be improved
- if the model could be extended.

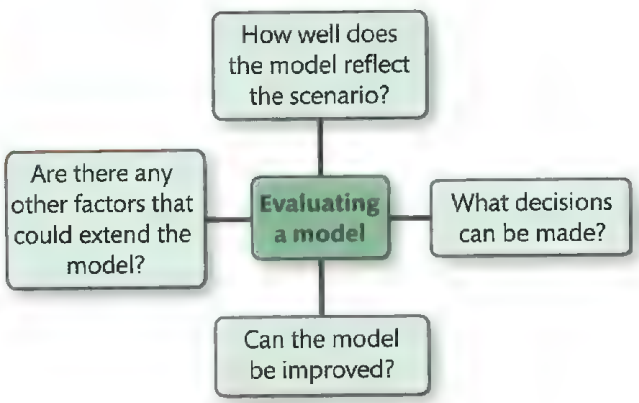


Figure 5.2: Evaluating models

How well the model reflects the scenario

When evaluating your data models, a good starting point is how well each of them mimics the given scenario. Carefully check each aspect of the scenario to confirm that it is included in the model. How well do the calculations match the scenario? Is it possible to check these calculations against values from the scenario?

Decisions that can be made using the model

The data model is there to support decision making, so what decisions does the model actually help you make? Are these decisions useful to the decision maker? Are there any decisions that the decision maker might need to make that are not supported by the model?

Whether the model can be improved

Evaluate how the model could be improved. Most models have potential for improvement, often at a cost in terms of time, expense or practicality. You need to be able to identify whether the model can be made better and weigh up the benefit or the extra value of these improvements against any extra time or costs these improvements would involve.

Other factors

Consider whether there are other factors that could be used to extend the model. There might be another data set that could be added to the model to improve the information it provides. Are there any other apps that could receive an export of data from your model to further extend its usefulness? An example of this would be a database using a query to join the export with other data to compile a report.

Tip

Try to think 'outside the box' to identify anything else that adds to the potential of the data model.

Documenting and justifying decisions

This documentation will summarise the situation (scenario) and will include your information sources, the factors you considered and the methods used to reach your decisions. It must include appropriate justification of those decisions.

Summarising the situation

You should start your documentation with a section that sets out the context for the report by outlining the scenario, the need for the data model and your role in this process.

Your documentation should have structure, that is, start with an introduction, then give the detail and end with a summary.

Identifying information sources

Include a section which identifies the sources of the data that you used for your model. Make sure that you state clearly where your data sets came from, the data they provide and how useful they are for the model.

Tip

A diagram could be used to communicate your sources of information.

Indicating factors considered

Your documentation needs to indicate the factors that you considered when planning your data model. This could include your current skills, the time within which you need to complete the model and any other constraints. You could include a sub-heading for each factor with a paragraph or two to explain why you thought the factor was relevant and any impact it had on your planning.

Methods used to reach decisions

You should include a section which outlines the different methods that you used to reach your decisions in your chosen model. For example, what calculations did you use to process the data into information? How did you show the information output from the model?

Justifying choices

Make sure that you justify:

- your choice of information sources
- the factors considered
- the methods used.

The justification could be included in each section and also summarised at the end of your report.

Reflect

Reflect on what you have covered in learning aim A and consider what you will need to record during the design, development and testing of your data model. This should include gathering feedback from others.

It is important to plan ahead and set yourself relevant targets and timescales. Reflecting on how well you achieved these after completing a project will help you plan for the next one.

Assessment practice 5.1

A.P1 A.P2 A.M1 A.D1

You have a job as a junior programmer in the head office of a retailer called W-Ready that specialises in selling weather-related goods at open-air events from a number of mobile shops, based in different parts of the country. W-Ready prides itself on selling goods that respond to the weather (for example umbrellas, waterproofs, wellies and hot soup at events where rain is expected or sunglasses, sunhats, ice creams and cold drinks during hot weather).

The purchasing manager, Ellie, wants to improve her decision making related to pre-ordering, so as to save the company money by enabling them to negotiate better prices from suppliers and reduce wastage and the need to store stock.

Ellie has requested that the programming team in which you work should produce a data model to help her to make purchasing decisions. She is not sure that a spreadsheet would be the right platform for this, so she has asked you for a report that evaluates the role of spreadsheet data modelling in the decision-making process.

Your report needs to include:

- the stages involved in the data modelling decision-making process
- explanations of how each stage would help focus on Ellie's requirements and maximise the usefulness of the model
- the features of spreadsheet software that can be used to support the decision-making process, including how they can be used in data modelling
- an analysis and evaluation of how spreadsheet software features would help support Ellie's decision-making.

Your report should be a comprehensive, detailed explanation of the stages involved in the decision-making process. It should consider how a systematic approach using valid information leads to informed decisions that can be justified. It should evaluate the use of advanced spreadsheet features and functions in data modelling to clearly show how this contributes to the decision-making process.

Plan

- What is the task? What am I being asked to do?
- What are the success criteria for this task?
- What am I learning? Why is this important?
- How will I approach the task? Are there any areas I think I may struggle with?
- How confident do I feel in my own ability to complete this task?
- Do I need clarification around anything?

Do

- Have I spent some time planning out my approach to the task at hand?
- Am I confident that I know what I am doing and what it is I should be achieving?
- What strategies am I employing? Are these right for the task? Are they working? If not, what do I need to change?
- I understand my thought process and why I have decided to approach the task in a particular way. I can explain this reasoning when asked.
- I can set milestones and evaluate my progress and success at these intervals.
- I understand when to consider and when to be decisive (reflection versus action).

Review

- I can explain what the task was.
- I can explain how I approached the task.
- I can say whether I met the task's criteria (ie whether I succeeded).

B Design a data model to meet client requirements

In this section, you will learn and practise the techniques needed to design your data model. You will start with the functional specification where you will define exactly what is required, and then progress to planning the spreadsheet that you will create to meet these client requirements. Finally, you will need to review your plans and adjust your designs to resolve any issues that you may find.

Functional specification

Your functional specification will be a document which needs to define everything that the data model needs to include so as to meet the client requirements. It should include

- the nature of the problem
- the functions that the model must perform
- the user interface
- any constraints (that is, limiting factors)
- the success criteria.

Nature of the problem

The nature of the problem makes a good introduction to your specification document as it puts the data model into context. The scenario should be summarised along with the users' needs and the client's broad expectations for the model (that is, what information they intend to get out of it).

Functions the model must perform

All spreadsheet functions that the model needs to perform must be listed in the specification along with enough detail to clearly define each of them. For example, whether the model needs to calculate averages or to count how many items are present in a data set.

User interface

You will need to include everything that is required from the **user interface** in the functional specification. This information should be taken from the scenario and user requirements.

Key term

User interface – the part of the data model that the user interacts with. This includes click-on buttons, other controls and any screen outputs that they can view, print or export.

Diagrams are not appropriate in the functional specification, because the specification should define and explain what is required. However, the design documentation (which will be described later) shows how you will implement your data model and, for this, diagrams are useful. In summary, the functional specification will specify what is needed from the user interface, but does not include the actual design for the user interface.

Constraints

Constraints need to be listed and explained in the functional specification. An obvious constraint to data modelling is the amount of time you have been allocated to produce your data model. Another will probably be the software available to you to implement your model. Make sure that you include all the constraints that you can identify in your functional specification.

Success criteria

This section defines what is needed to confirm that the data model is acceptable and complete. If you are employed as a data modeller who completes data models for clients, meeting the success criteria usually triggers the payment for a job because it indicates that it is finished.

An often overlooked but important part of the specification is the definition of the success criteria at the outset. This is very good practice as all interested parties are then clear on exactly what the data model must deliver before the model is produced. Identifying success criteria means understanding what benefits a business needs to get out of the data model (that is, the information that they require) and agreeing on client requirements.

PAUSE POINT List all the constraints that you might expect for a data modelling project.

Hint: Think about anything that might hinder achieving a perfect data model.

Extend: How would an IT professional, working for an organisation, deal with these constraints?

Spreadsheet model design

Your designs for your data model will show how you are intending the spreadsheet model to look and work. These designs will help you to communicate to everyone involved in the project how the model will look and function. They will also be useful for you, to make sure that you are clear about what you are required to produce.

Worksheet structure diagrams

You will need to produce worksheet structure diagrams to show:

- the layout and presentation
- data processing
- data entry and validation
- navigation
- the output of your planned data model.

Layout and presentation

Your layout diagrams should show how you will structure the spreadsheet. They will show where the inputs and outputs are to be located as well as where the spreadsheet will keep any data needed by your model. All the worksheet pages should be shown here along with details of what each of them will contain.

Presentation diagrams should show how you plan to communicate the information that the data model produces. It should include your choice of chart type(s) and where they will be held in the spreadsheet.

Processing

The processing section diagrams need to outline the calculations (formulae and functions) and other actions, such as copying data, which the spreadsheet will carry out. These could be simple **block diagrams** to show where your calculations will be, along with what they do.

Any macro coding will require **program flowcharts** or **Jackson Structured Programming (JSP) structure diagrams** to show how your code will work. Macros are the code you write to program a spreadsheet to carry out certain actions on command. Macro code for Excel® is called Visual Basic® for Applications (VBA) and is very similar to the Visual Basic .NET (VB.NET) programming language.

Data entry and validation

You will need diagram(s) to show the data entry and validation for your data model. These will clearly identify the cells or forms where data is typed. You could annotate the diagram(s) to explain any validation you plan so as to minimise mistyping and other data entry errors.

Key terms

Block diagram – a simple diagram that uses rectangles to represent the parts of a system (such as a section of program code) to show how it is structured. These diagrams communicate how the parts fit together without the need for unnecessary detail.

Program flowchart – a diagram using shapes and lines to show where decisions (branches) are taken, when inputs and outputs are made and where the code can end. These diagrams show the routes that can be taken through program code.

JSP structure diagrams – Jackson Structured Programming (JSP) structure diagrams are a type of a block diagram used to show how programs are structured. They have lines joining the blocks to show the sequence, decisions and iterations.

Navigation

Your navigation diagram(s) will probably be quite simple and use block diagrams to represent the worksheets in your planned spreadsheet and arrows to show how the user can move from one part of the data model to another. The actions needed for these navigations should be included in these diagrams (for example, clicking on a button, pressing a key on the keyboard or clicking on other parts of the spreadsheet).

Output

The outputs from your data model are important because they convey the meaning derived from your data model, so they need to be carefully planned. These diagrams should clearly show what information is being output along with where they can be found in the data model.

You can produce mock-ups of any charts that you are planning, including annotations, to clarify any features you will include in them. Carefully choose the chart types that you think best convey the meaning(s) of your information. Block diagrams with annotations can be used to communicate how you have planned the printed or on-screen outputs from your data model.

Test plans

You will need to produce a test plan for the data model that shows your choice of test data, the purpose of each test and the results you expect to get from each test.

You can use your functional specification to structure the test plan, ensuring that the tests cover all aspects of your data model. Table 5.1 on the following page shows an example of the headings that you could use in your test log to record outcomes of testing.

Table 5.1: Test log

Test number	Test description	Test data	Expected result	Actual result	Test passed	Action taken	Re-test number

Test data

The most effective testing uses carefully chosen test data to confirm each calculation, produces the correct answers, finds conditions that respond as expected and finds that bad data does not break the model.

Purpose of the test

Every test must have a purpose that should be recorded in the 'Test description' column of your test log. Knowing the purpose of the test is especially important to help you focus your efforts when testing the model. If a test has no purpose, there is no point in doing it.

Expected result

The expected result of every test needs to be calculated or predicted before the test is run. Test data results can be calculated separately before using your model so that, when they are put through your spreadsheet, you can recognise whether each test has been passed or not.

If a test has not been passed, you should first check your pre-test calculation to confirm the expected result. The data model will then need some adjustment and the test should be re-run to confirm that the correction has solved the problem. All of this needs to be recorded in the test log.

Link

For more on testing see section Testing the data model solution.

Reviewing and refining data model designs

You should now have a design for your data model which is the result of a lot of research, thought and planning. In a business, you would usually carry out this work in a team. Collaborative working like this usually brings positive contributions from each person in the team so as to produce the best design for your data model.

Once you have a design for your data model, you would share this with the client, other stakeholders and any other potential users. You would then work as a team to improve the quality, effectiveness and appropriateness of your designs based on their feedback.

Gathering feedback from clients and potential users

You will find it particularly useful to gather feedback from both the client and potential users on the extent to which your design meets their requirements. Client feedback is a good starting point for reviewing and refining your design, simply because they know what they want (because they defined the requirements) and are well placed to communicate with you on how well your planning is meeting those requirements. Potential users should be able to give feedback on how easy or difficult they think your model would be to actually use in the workplace.

Communicating with clients

The methods of communication that you use with clients will need to be appropriate for each situation. Emails are quick and easy and allow you to attach the design documents for the client to look at when it suits them. Similarly, emails are an easy way for the client to respond. However, it is best to avoid communicating complex issues or problems via email as this can lead to misunderstandings. It is better to use email to arrange a meeting to deal with such issues. Verbal communication, especially in a meeting, can be very effective as issues which are raised can be discussed and agreed effectively, with any confusions being ironed out on the spot. You could send an email to confirm agreed actions after a meeting.

Meetings

Meetings can be the most powerful type of communication, although more time consuming due to the need to set up the meeting, hold it and to document it afterwards. Scheduling a meeting needs to be carefully arranged to fit in with the diaries of all the attendees, but also needs to be held before any issues it will raise might become critical.

Risks, Actions, Issues and Decisions (RAID) logs can be used to document the meeting. A copy should be sent to each attendee after the meeting to confirm what was discussed and to ensure that agreed actions and outcomes are accurately understood.

Key term

RAID logs– Risks, Actions, Issues and Decisions (RAID) logs are often used in project work to record things that could go wrong (Risks), what needs doing (Actions), problems (Issues) and the agreements made (Decisions).

Timescales

The timescales for implementing your data model should be agreed by all concerned in the data modelling project (that is the client and the data modellers) and will depend on what is possible and practical. There might be a need to adjust these timescales if you encounter problems during the planning or implementation stages. In this case, all the project stakeholders should be informed and new timescales should be agreed.

Refining ideas and solutions

The feedback you receive from clients and potential users, when they review the design of the data model, will almost certainly result in the need to refine your ideas and solutions to produce a better, more useful and realistic plan for the data model.

Review and feedback

Changes will need to be made to your planning documentation by updating the design specification, based on the reviews and feedback.

Link

For more on review and feedback see Reviewing and refining the data model solution.

Reflect

You will have reviewed your data model using feedback from professionals on the quality of the data model and its suitability to meet the original design requirements. The views of other people are very valuable as they can bring a fresh and honest opinion on the real worth and usefulness of your work.

Think about how you would respond positively to negative feedback about your data model from a client or potential users.

PAUSE POINT How will you review your designs with others, to identify improvements?

Create a list of all the people who you could ask to review your data model designs, along with what each person would be best at identifying.

Extend How would you get your reviewers to communicate their findings to you?

C Develop a data model to meet client requirements

Having designed and planned your data model, you can now begin to build up your spreadsheet skills by finding out how to use Excel®. Once you have developed your model, it will need to be thoroughly tested and any problems will need to be fixed before being reviewed and undergoing further refinement. This project work will provide great opportunities to improve the quality of your written documentation and other communication skills.

Developing a data model solution

You will need to develop and practise your skills at using the Excel® data processing features. You will need to be able to create formulae and functions and forms for data entry which have the required layout to aid presentation and which meet the output requirements.

Processing features and requirements

Excel® offers many useful processing features to enable you to fulfil your processing requirements. These features are discussed in this section.

Formulae

Excel® makes it very easy to enter simple arithmetic formulae to add, subtract, divide and multiply.

All you need to do is to type =, select one of the cells for the formula, type the arithmetic sign, select the other cell in the formula then use the <Enter> key, for example = A3 * B3.

Link

For more on formulae see the Using formulae and functions section.

A formula can have more than two **arguments** (terms), for example = A3 * B3 / C3, and can mix cell references with numbers, for example = 2 * B3 / C3.

When you have more than two arguments, you might find the result of a formula is different from what you expect. This will be due to **BIDMAS (or BODMAS)**, otherwise known as the precedence of operators.

Discussion

What is the answer to 3 + 4 * 5? What answer does Excel® give to this simple formula? Why?

Key terms

Argument – part of a calculation (formula or function) that represents a number or value. An argument can be a cell reference such as A2 or a value such as a number, say, 12, or it can be text, say, Richard.

BIDMAS (or BODMAS) – stands for Brackets, Indices/Orders, Divide, Multiply, Add, Subtract and is used to remind us of the order in which a formula is worked out (otherwise known as the precedence of operators).

Worked example: BIDMAS

From BIDMAS, we can see multiplication is worked out before addition and that brackets are worked out first before multiplication.

4 + 5 * 6 works out to be 34 because 5 is multiplied by 6 before 4 is added.

If you want the answer to be 54, the formula needs brackets, (4 + 5) * 6 so that 4 and 5 are added together first (because they are in brackets) before being multiplying by 6.

Relative and absolute cell referencing

The formulae used so far in this unit have all been relative, for example A3, which are easy to enter and good when you want to copy and paste a formula, because the cell references adjust in the pasted formulae. In Figure 5.3, the formula was entered into cell D3 as =C3 * B3 and, when it was copied and pasted into cells D4, D5 and D6, the formula adjusted so that each time it multiplied the two cells to the left.

These are known as relative cell references because a copied formula changes to keep the relative positions. In the example, the second cell to the left gets multiplied by the first cell to the left.

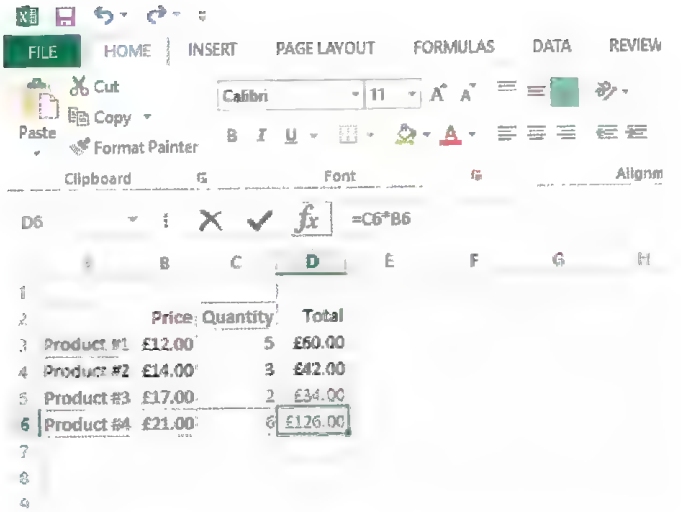


Figure 5.3: Relative formula copy

For many formulae, relative addressing is perfect and good for copy and paste. However, other formulae need to also use absolute cell referencing to prevent some parts of the formula from changing when using copy and paste. In Figure 5.4, the formula was entered into cell E3 as =D3 * $E$1 and, when it was copied and pasted into cells D4, D5 and D6, the absolute part of the formula, $E$1, stayed the same so that, each time, this was multiplied by the cell to the left.

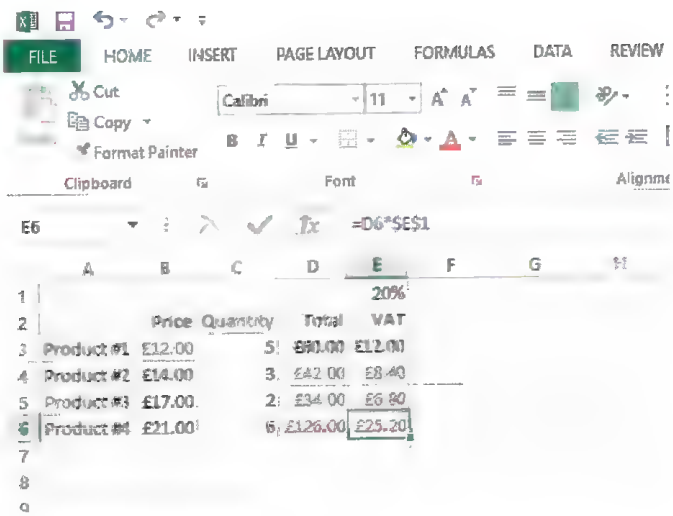


Figure 5.4: Absolute formula copy

These are known as absolute cell references because they do not change in a copied formula.

Closer examination of an absolute cell reference shows two $ signs, one before the column letter, the other before the row number. These can be used for more control over copying. For example, if $E1 is copied and pasted, the column letter (E) will stay the same, while the row number (1) will change as a relative cell reference.

Tip

In Excel you can use the <F4> key to put the $ signs into a relative cell reference to make it absolute.

Functions

A function is a complex calculation that is built into Excel®. Every function has a name and usually needs arguments inside brackets following the name.

To use a function you need to type = followed by the first part of the function name. You can then select the function you want, making sure that the terms (arguments) needed by the function are there, and then press the <Enter> key: for example, = AVERAGE(A1:C4).

Link

For more on functions, see the Using formulae and functions section.

Many IT professionals use the prompt that Excel® shows under a function as it is being entered to help remind them what each argument does. The arguments in a function are separated by commas. When a comma is typed, the prompt changes to describe the current argument.

There are around 400 functions in Excel ®, so it is easy to carry out calculations for almost any type of data model. Excel® also allows programmers to use macro Visual Basic® for Applications (VBA) code to add their own functions to this collection. Figure 5.5 shows some of the general functions available in Excel® being applied and Table 5.2 lists some general functions in Excel® and examples of how they can be used.

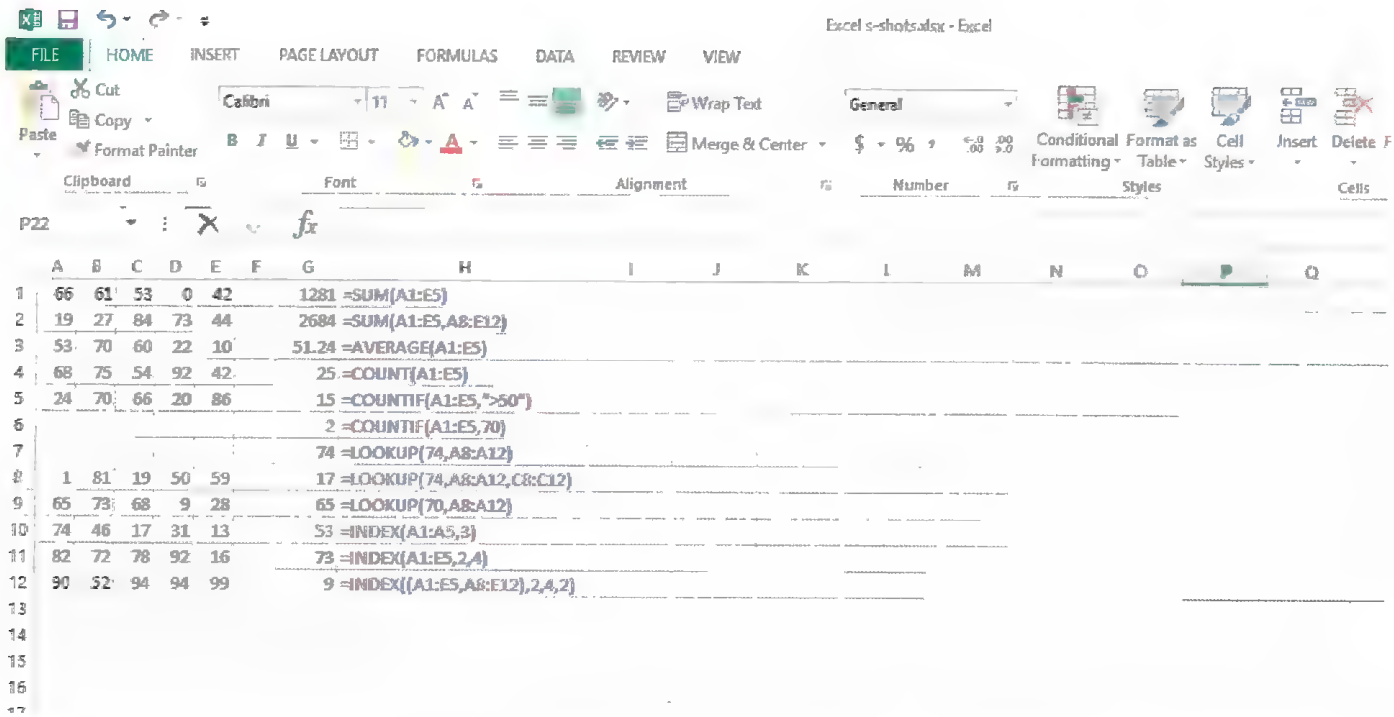


Figure 5.5: Examples of general functions in Excel®

Table 5.2: General functions in Excel®

Function	Arguments	Why it is used	Example of how it can be used
=SUM(Number1 [,Number2],...)	Number1 is a cell or number or range of cells Number2 is an optional further cell, number or range of cells	To add the arguments together.	• =SUM(A1:E5) • To add all the numbers in the cells from A1 to E5. • =SUM(A1:E5,A8:E12) • To add all numbers in the cells from A1 to E5 and A8 to E12.
=AVERAGE(Number1 [,Number2], ...)	Number1 is a cell or number or range of cells Number2 is an optional further cell, number or range of cells	To show the average of the arguments.	• =AVERAGE(A1:E5) • To average all the numbers in the cells from A1 to E5.
=COUNT(Range1 [,Range2], ...)	Range1 is a cell or number or range of cells Range2 is an optional further cell, number or range of cells	To count how many numbers are in the range(s).	• =COUNT(A1:E5) • To count how many numbers are in the cells from A1 to E5.
=COUNTIF(Range, Criteria)	Range is the cells that are searched to count how many of the criteria are found Criteria is what is counted	To count how many cells in the range meet set criteria.	• =COUNTIF(A1:E5,'>50') • To count how many numbers in cells A1 to E5 are bigger than 50. • =COUNTIF(A1:E5,70) • To count how many 70s are in the cells A1 to E5.
=LOOKUP(Value, Lookup [,Result])	Value is what is looked for Lookup is where to search and must be in order, from low to high Result is an optional range for what gets shown by this function and must be the same size as the Lookup range	To find a value and show either the best match or, optionally, whatever is in the Result cell corresponding to where the value is found in the Lookup range. See Figure 5.6 for examples of VLOOKUP() and HLOOKUP() functions.	• =LOOKUP(74,A8:A12) • To find 74 in the range A8 to A12, as there is an exact match, 74 shown. • =LOOKUP(74,A8:A12,C8:C12) • To find 74 in the range A8 to A12 and show the match from C8 to C12. As there is an exact match, 17 is shown. • =LOOKUP(70,A8:A12) • To find 70 in the range A8 to A12, as there is no exact match, 65 is shown.
=INDEX(Ref, Row [,Col] [,Area])	Ref is the cell range and needs brackets if more than 1 range is used Row is the number of the cell in the range which is to be shown Col is optional for when the cell range is a rectangle Area is optional for when there is more than 1 range to select which area is to be used	To show what is in a cell in a range.	• =INDEX(A1:A5,3) • Shows 53 as that is inside cell A3 as row 3 is the third row of the range A1 to A5). • =INDEX(A1:E5,2,4) • Shows 73 as that is inside cell D2 as row 2 is the second row of the range A1 to E5) and col 4 is D (fourth column of A1 to E5). • =INDEX((A1:E5,A8:E12),2,4,2) • Will show 9 as that is inside cell D9, row 9 is the second row of the second range A8 to E12) and the fourth column is D.

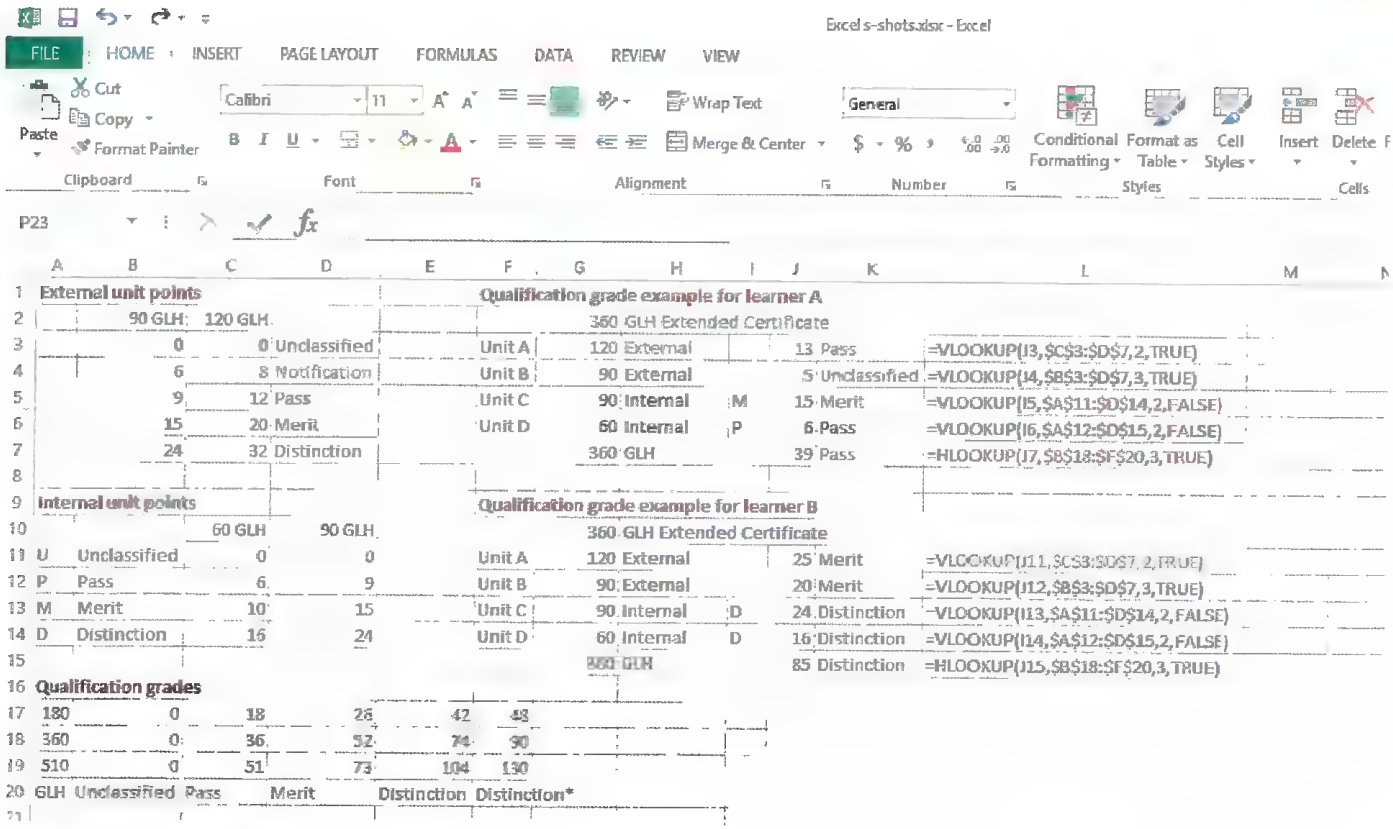


Figure 5.6: VLOOKUP() and HLOOKUP() functions

Table 5.3: Other lookup functions (see Figure 5.6: VLOOKUP() and HLOOKUP() functions, for examples)

Function	Arguments	Why it is used	Example of how it can be used
=VLOOKUP(Value, Table, Col [,Lookup])	Value is what you are looking for in the table Table is where you are looking for the value Col is the column in the table for the result Lookup is optional, choose between True (approximate match for numbers) and False (for an exact match)	To find something in a range of cells to bring back a result matching the find. V (vertical) lookup is used when the left column of the table holds what you are looking for. This is a good alternative to nesting lots of IF() functions as it is much easier to use.	• =VLOOKUP(J3,$C$3:$D$7,2,TRUE) • Shows 'Pass' as that is inside cell D5 as row 5 is the row where 13 (from J3) is found in the first column of range C3 to D7 and column 2 in the table is D. • The TRUE in the function is for approximate match which allows 13 (from J3) to be found. • The range C3 to D7 is entered as $C$3:$D$7 so the function can be copied and still point to the same table. • =VLOOKUP(I5,$A$11:$D$14,2,FALSE) • Shows 'Merit' as that is inside cell B13 as row 13 is the row where 'M' (from I5) is found in first column of range A11 to D14 and column 2 in the table is B. • The FALSE in the function is for exact match as only 'M' (from I5) is to be found. • The range A11 to D14 is entered as $A$11:$D$14 so the function can be copied and still point to the same table.
=HLOOKUP(Value, Table, Row [,Lookup])	Value is what you are looking for in the table Table is where you are looking for the value Row is the row in the table for the result Lookup is optional, choose between True (approximate match for numbers) and False (for an exact match)	To find something in a range of cells to bring back a result matching the find. H (horizontal) lookup is used when the top row of the table holds what you are looking for. This is a good alternative to nesting lots of IF() functions as it is much easier to use.	• =HLOOKUP(J7,$B$18:$F$20,3,TRUE) • Shows 'Pass' as that is inside cell C20 as column C is where 39 (from J7) is found in the first row of range B18 to F20 and row 3 in the table is 20. • The TRUE in the function is for approximate match which allows 39 (from J7) to be found. • The range B18 to F20 is entered as $B$18:$F$20 so the function can be copied and still point to the same table.

Discussion

Figure 5.6 (VLOOKUP() and HLOOKUP() functions) shows some examples of how these functions are used. Consider the following discussion points which all relate to this spreadsheet.

- Why are the functions in L3, L4, L7 looking for approximate matches?
- Why are the functions in L5, L6 looking for exact matches?
- What function could be used in cell J5 to look up the points for this 90 GLH (Guided Learning Hours) unit?
- What function could be used in cell J6 to look up the points for this 60 GLH unit?
- Can you write a function for cell J5 which can then be copied to cells J6, J13, J14 to give accurate results?
- Can you write a function for cell L3 which can then be copied to cells L4-L6, L11-L14 to give accurate results which cater for External/Internal and the number of GLH? (You could insert column(s) to provide interim calculations to simplify your function.)
- What function would you use in cell J7? Can this be copied to cell J15?
- What function would you use in cell G7? Can this be copied to cell G15?
- How could you apply conditional formatting to cell G15 to validate the unit GLH numbers?

Logical functions

These functions use logic to decide upon a choice of outcomes. The logic is anything that results in a true or false result, for example comparing two values to find which is greater (larger) than the other. See Figure 5.7 for examples of logical functions.

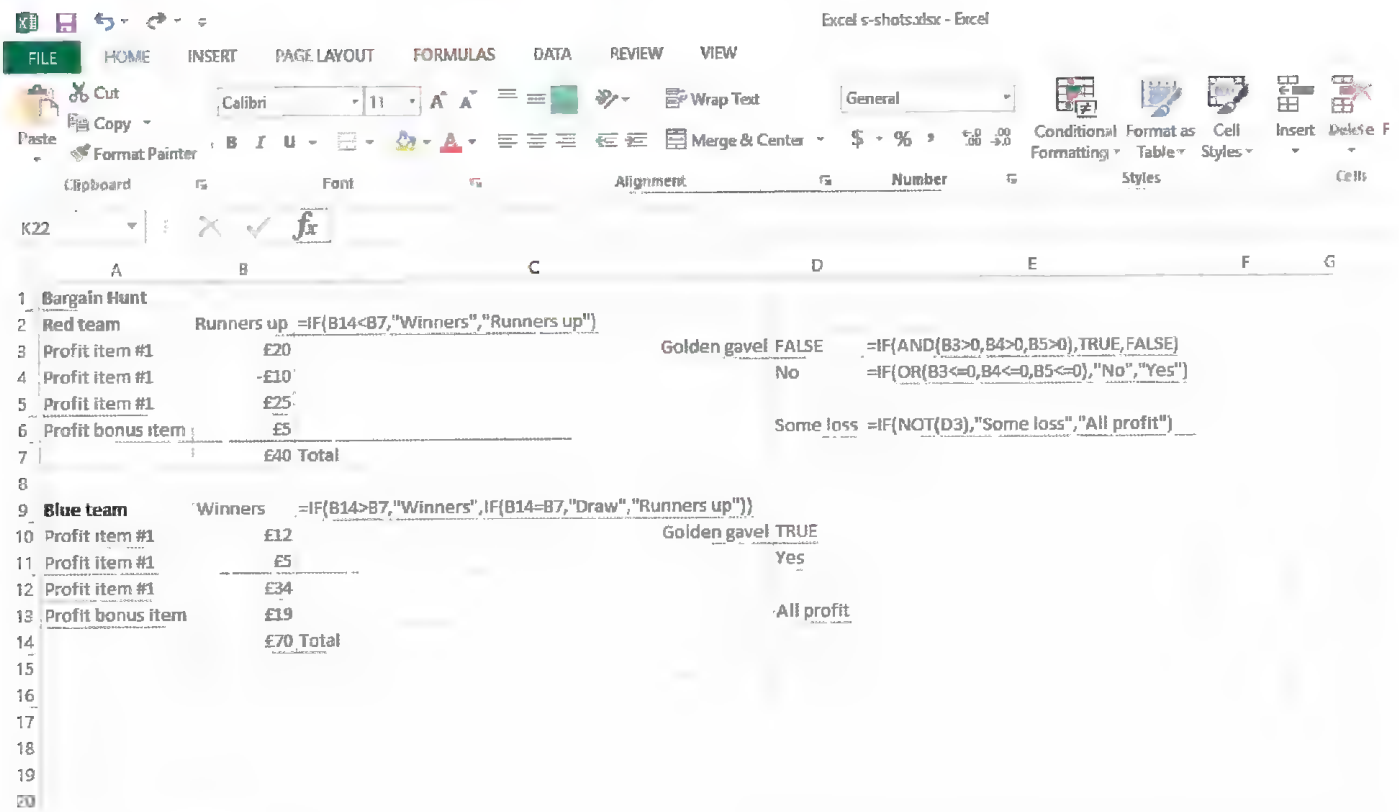


Figure 5.7: Logical functions

Table 5.4: Logical functions (see Figure 5.7 Logical functions, for examples)

Function	Arguments	Why it is used	Example of how it can be used
=IF(Condition, IfTrue [,IfFalse])	Condition is anything that can be worked out as true or false IfTrue shows if condition is true IfFalse it is optional what to show if the condition is false	When a cell needs to show one of two possible results.	• =IF(B14<B7,"Winners","Runners up") • To test if the contents of cell B14 is less than what is in cell B7 to show "Winners" if true or "Runners up" if false.
=NOT(Condition)	Condition is anything that can be worked out as true or false	To flip (invert) a true condition to false or a false condition to true.	• =NOT(D3) • To invert the contents of cell D3. • D3 is showing false, so the example here returns a true. • This is used as the condition of the IF() function in cell E6.
=AND(Condition1 [, Condition2], ...)	Condition1 is anything that can be worked out as true or false Condition2 is anything that can be worked out as true or false	To show true if each of one or more conditions are true.	• -AND(B3>0,B4>0,B5>0) • To test if the contents of cells B3, B4, B5 are each more than zero (0), returns true if all of these are above zero. • This is used as the condition of the IF() function in cell E3. The = is not needed here as the AND() is inside the IF() function.
=OR(Condition1 [, Condition2], ...)	Condition1 is anything that can be worked out as true or false Condition2 is anything that can be worked out as true or false	To show true if any of one or more conditions are true.	• =OR(B3<=0,B4<=0,B5<=0) • To test if the contents of cells B3, B4, B5 are each less than or equal to zero (0), returns true if any of these are zero or below. • This is used as the condition of the IF() function in cell E4. The = is not needed here as the AND() is inside the IF() function.

Nested IF functions

Nesting means that you have a function inside another function. The inside function is worked out first so this result can then be used in the main function. Nested IF() functions can be used for a cell showing one of three possible outcomes. See Figure 5.7 and Table 5.4 for an example of this where cell C9 contains an IF() function to answer which team won by comparing the total of the Blue team (B14) to find if it is greater than the total of the Red team (B7). If true, "Winners" is shown, If not true, the nested IF() function responds to a draw (B14=B7), showing "Draw" if true or "Runners up" if false.

Figure 5.7 and Table 5.4 show examples of how NOT(), AND() and OR() functions can be nested inside IF() functions. Most functions can be nested inside another function. The result given by the nested function is used by the main function.

Tip

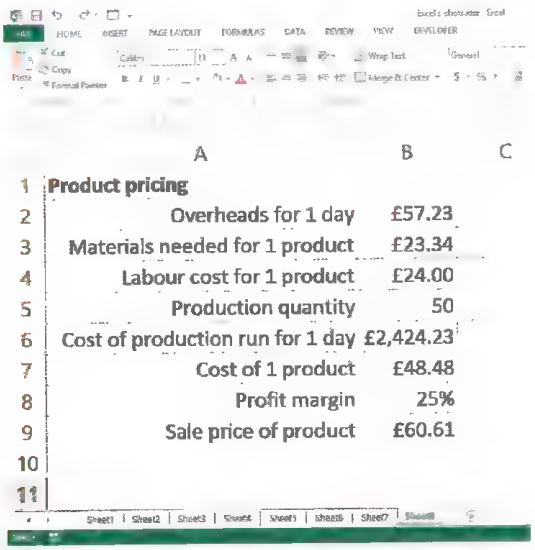
Nested IF functions can be used for more than three possible outcomes, but this is not recommended as it is easy to make mistakes and hard to solve any errors in the result. It is much better to use VLOOKUP() or HLOOKUP() for a cell that needs four or more possible outcomes.

Goal seek

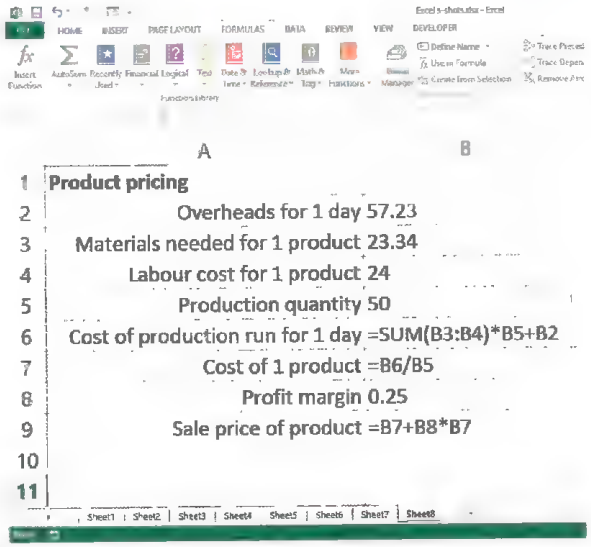
Goal seek is a feature of Excel® that allows you to select a cell holding a calculation then choose the answer you want from this cell by selecting another cell that can be changed to provide the answer. Excel® will then work out what needs to be in the changeable cell to provide the answer.

Step by step: Using Goal Seek

1 This step by step shows how to use Goal Seek to adjust the number in a cell to produce a required result in a calculated cell. The example here is a very simple data model used to explore the costs of a manufactured product. Firstly, set up the data model using calculations and functions as shown here.

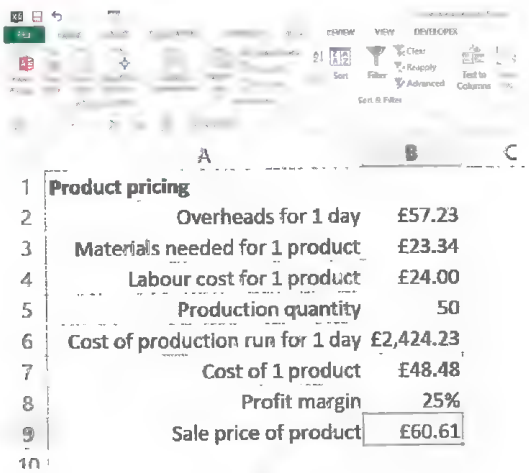


	A	B	C
1	**Product pricing**		
2	Overheads for 1 day	£57.23	
3	Materials needed for 1 product	£23.34	
4	Labour cost for 1 product	£24.00	
5	Production quantity	50	
6	Cost of production run for 1 day	£2,424.23	
7	Cost of 1 product	£48.48	
8	Profit margin	25%	
9	Sale price of product	£60.61	
10			
11			



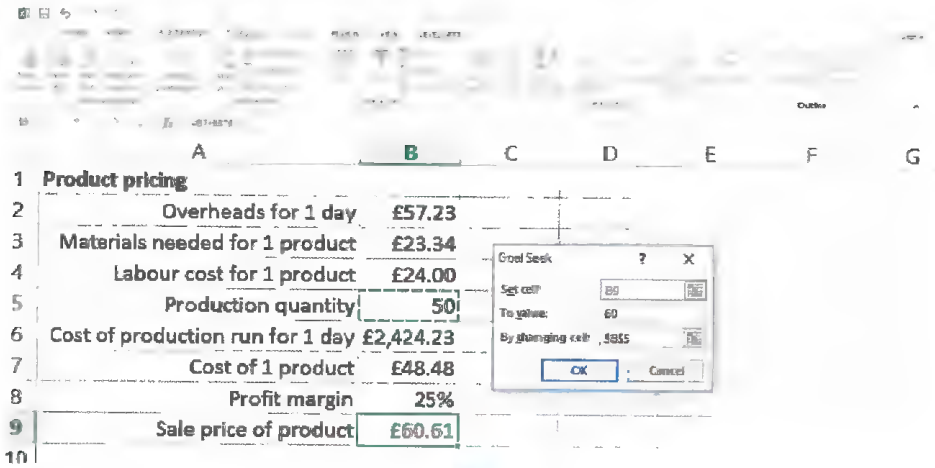
	A	B
1	**Product pricing**	
2	Overheads for 1 day	57.23
3	Materials needed for 1 product	23.34
4	Labour cost for 1 product	24
5	Production quantity	50
6	Cost of production run for 1 day	=SUM(B3:B4)*B5+B2
7	Cost of 1 product	=B6/B5
8	Profit margin	0.25
9	Sale price of product	=B7+B8*B7
10		
11		

2 Market research has shown that the maximum sale price could be £60, so this data model is to be used to find out how many products need to be produced each day to meet that price. Cell B9 is selected then the Goal Seek option is taken under the What if Analysis drop-down menu in the Data toolbar.



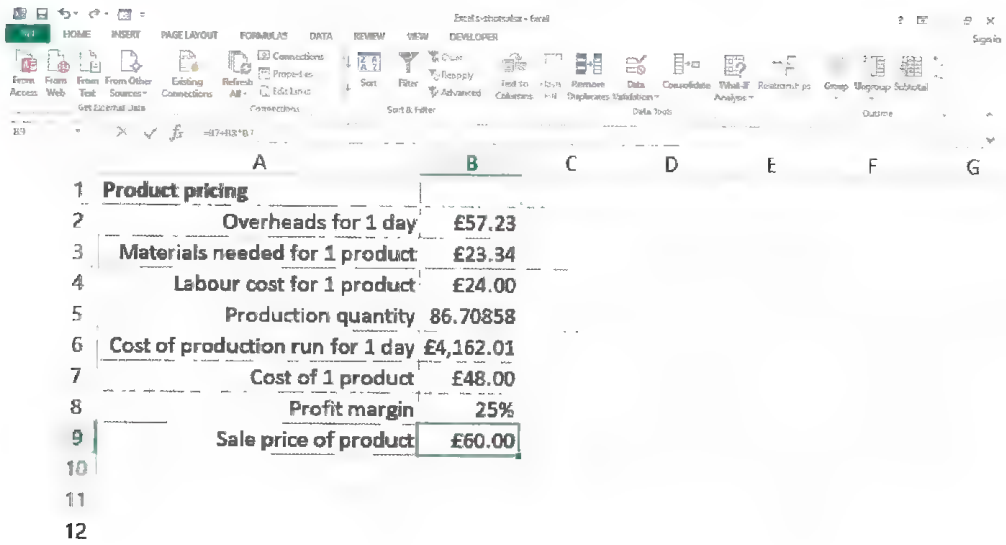
	A	B	C
1	**Product pricing**		
2	Overheads for 1 day	£57.23	
3	Materials needed for 1 product	£23.34	
4	Labour cost for 1 product	£24.00	
5	Production quantity	50	
6	Cost of production run for 1 day	£2,424.23	
7	Cost of 1 product	£48.48	
8	Profit margin	25%	
9	Sale price of product	£60.61	
10			

The Goal Seek dialog box is completed to set B9 to 60 by changing B5.



	A	B
1	**Product pricing**	
2	Overheads for 1 day	£57.23
3	Materials needed for 1 product	£23.34
4	Labour cost for 1 product	£24.00
5	Production quantity	50
6	Cost of production run for 1 day	£2,424.23
7	Cost of 1 product	£48.48
8	Profit margin	25%
9	Sale price of product	£60.61
10		

The OK button is used to run the Goal Seek resulting in the data model showing the production run size to meet the price of £60. The target production quantity is shown with a decimal part, so would need to be rounded up to a whole number for a practical run size of 87 products per day.



	A	B
1	**Product pricing**	
2	Overheads for 1 day	£57.23
3	Materials needed for 1 product	£23.34
4	Labour cost for 1 product	£24.00
5	Production quantity	86.70858
6	Cost of production run for 1 day	£4,162.01
7	Cost of 1 product	£48.00
8	Profit margin	25%
9	Sale price of product	£60.00
10		
11		
12		

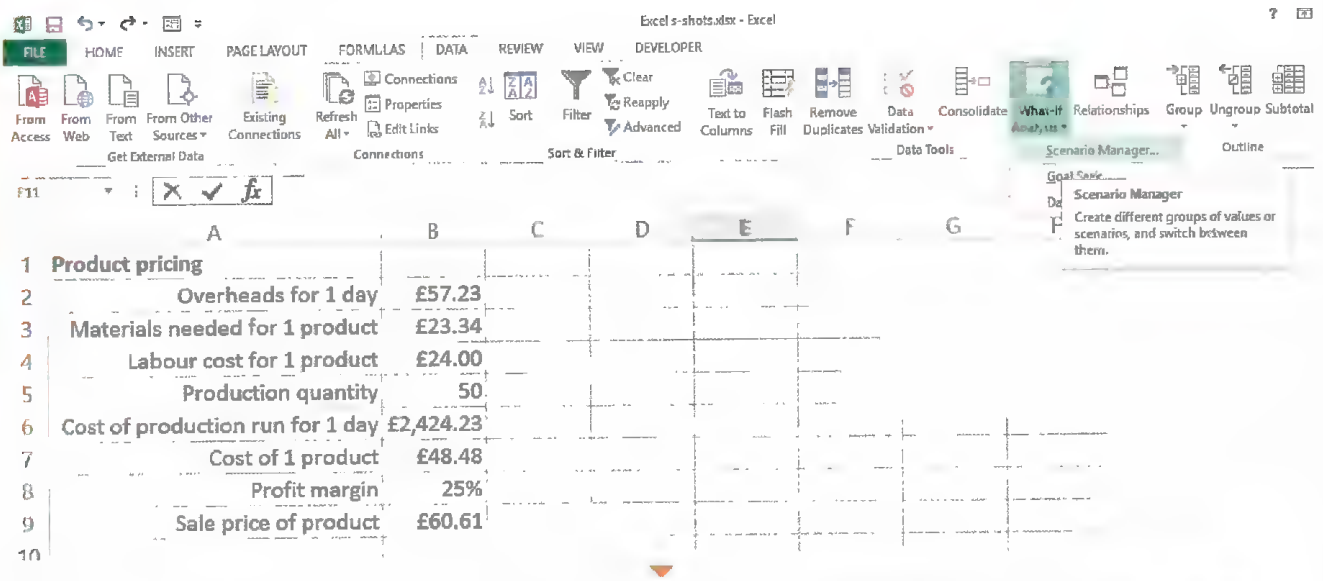
What if

What if is a feature of data models that allows you to explore different possibilities by changing numbers in cells that are used by the formulae and functions of the spreadsheet data model. As the calculations are almost instantaneous, it is very easy to explore options such as: 'What if we increase the cost of a product?' or 'What if our sales decrease by 10 per cent?'

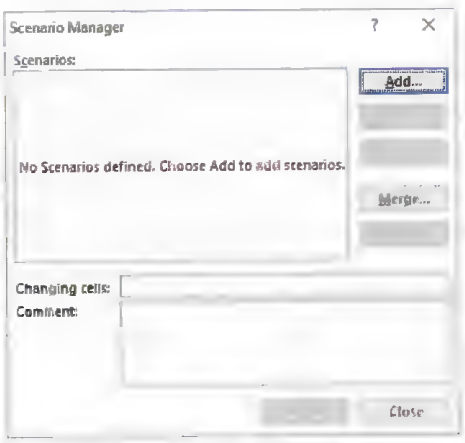
Good data modelling design will group together the cells used for key numbers used in the model and also have the key outputs close so it is easy to see the outcomes from different inputs without having to scroll to other parts of the spreadsheet. These What if explorations can be made by simply typing over existing values in a data model's cells. Excel® also offers some tools under the What if Analysis drop-down menu in the Data toolbar which include Scenario Manager, Goal Seek and Data Table.

Step by step: Using What if scenarios

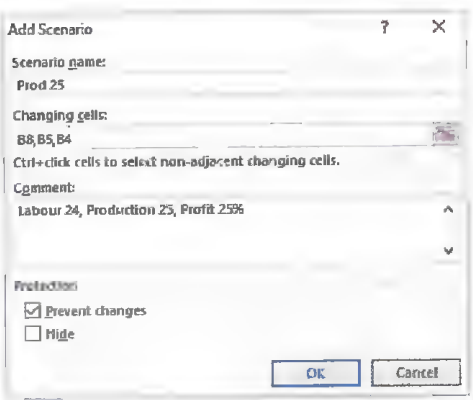
1 This step by step shows how to set up a What if scenario using the same data model shown in the Using Goal Seek step by step. Cell B8 is selected and then the Scenario Manager option is taken under the What if Analysis drop-down menu in the Data toolbar.



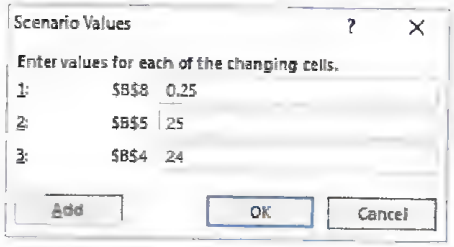
2 Once you have the Scenario Manager dialog box, use the Add ... button to add a scenario.



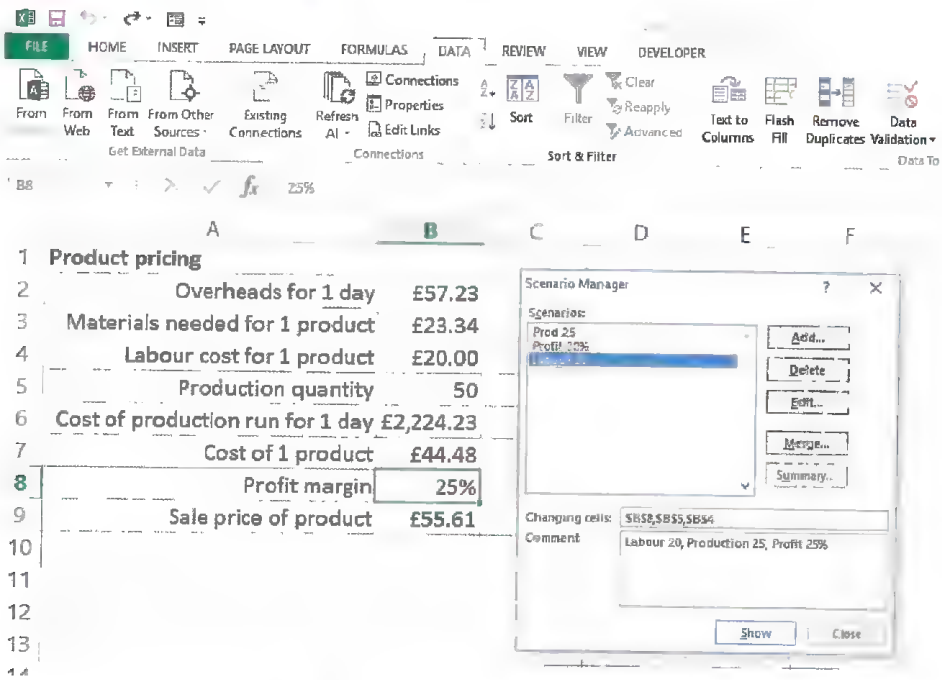
3 The scenario can be given a name so that it can be easily selected when completed. The cells that can change are selected. They will all have values, not formulae. The Comment defaults to who is using the spreadsheet and date, but can be changed to summarise the values used in his scenario.



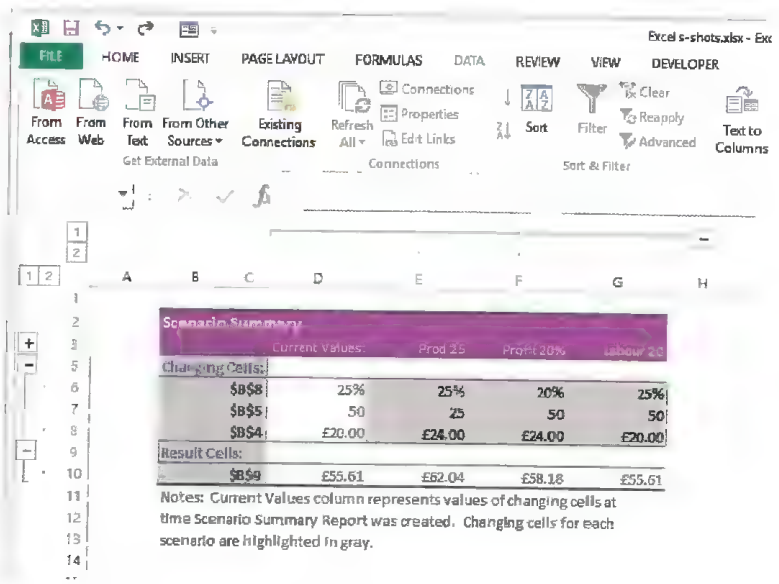
4 Using the OK button moves from the Add Scenario dialog on to the Scenario Values dialog so that the values to be used in the selected cells for this scenario can be set. Press the OK button once this is complete. More scenarios can be added with different values which are of interest to the model.



5 Any of your scenarios can then be selected from the Scenario Manager. Press the Show button to set the spreadsheet cells to that scenario's values.



6 The Scenario Manager has a Summary button which can create a new worksheet with a scenario summary. The summary has controls around it to control how much of the summary is shown.



Data manipulation

Data manipulation is a feature of spreadsheets which lets you see the data in different ways, such as:

- sorting into a different order
- grouping to bring similar or related items together
- filtering to only see the items you are interested in
- pivoting data to rearrange the rows and columns.

Sorting

Sorting a range of cells is useful in many ways. You might want to order items showing the cheapest at the top and the most expensive at the bottom, or list names ordered by surname or by first name. Sorting will also show duplicates as they would be next to each other.

A two-level sort will group the items together on the first-level sort. For example, this could be by the first level, department, then by the second level, surnames in alphabetical order.

Multilevel sorting can have more than two levels. For example, a three-level sort could group staff members by (1) location then by (2) department then by (3) surname. To do this, select the data and then use the Sort button in the Data toolbar to bring up the Sort dialog box, as shown in Figure 5.8.

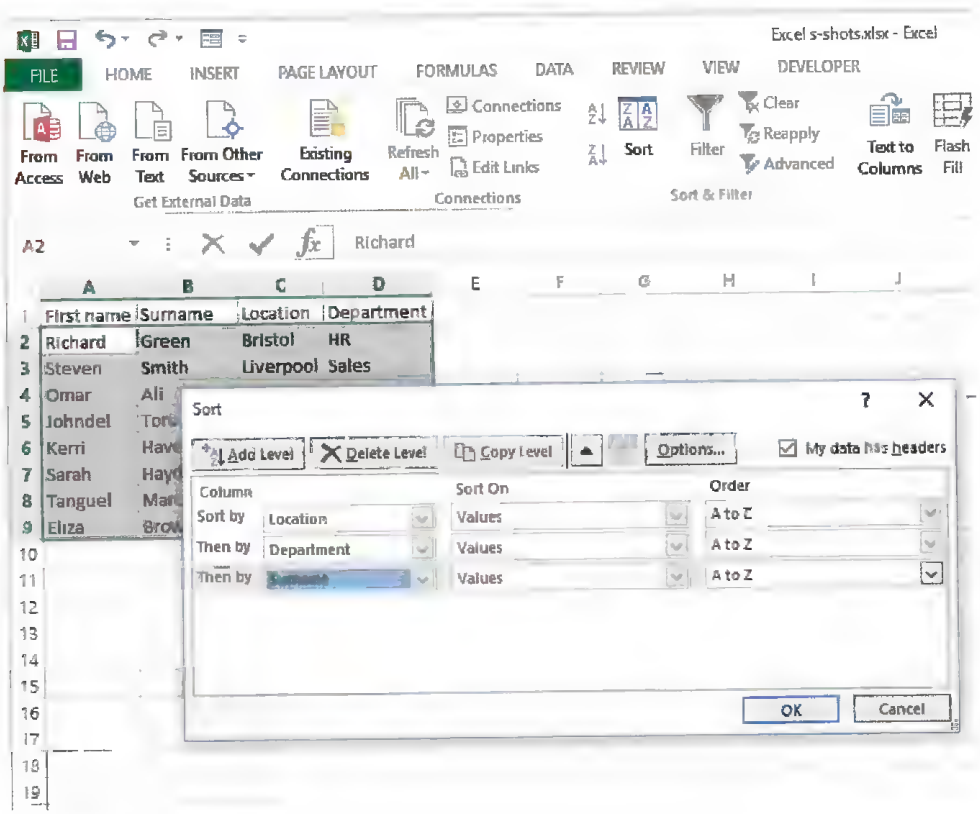


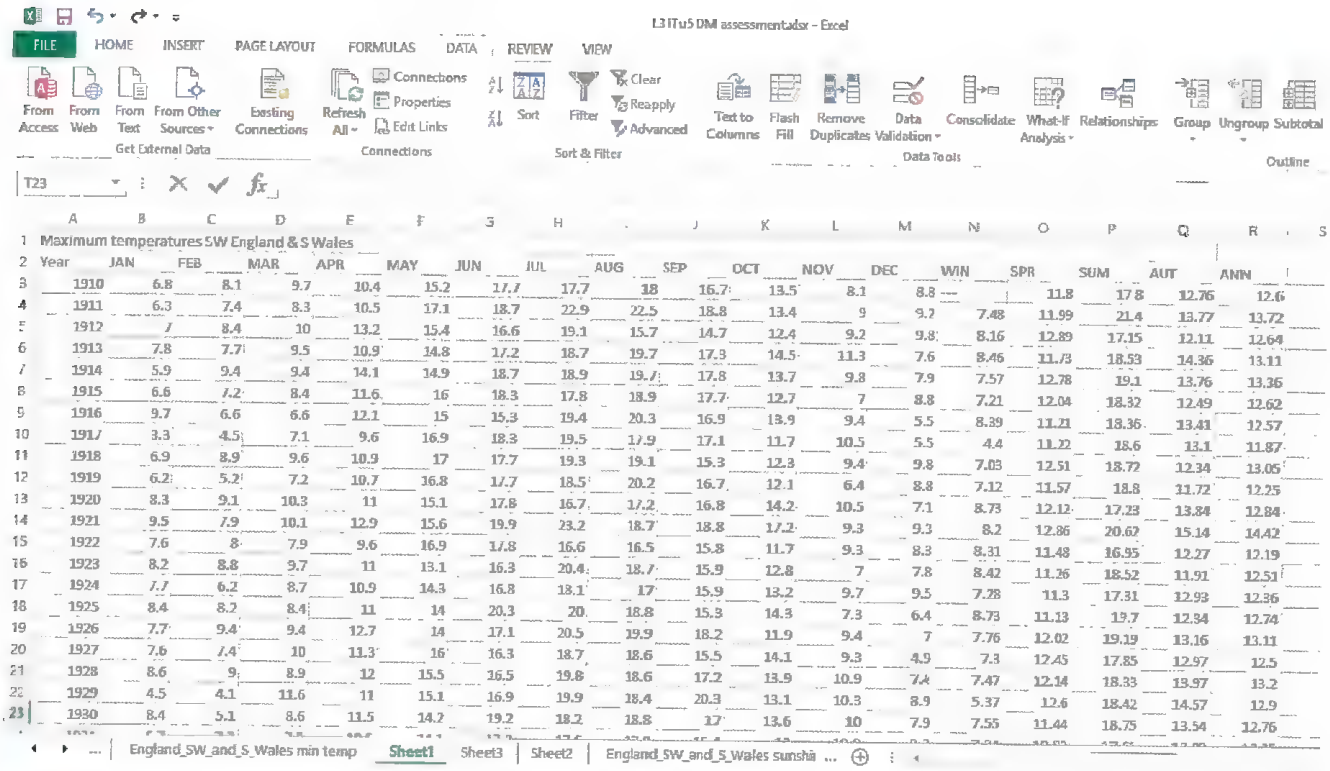
Figure 5.8: A three-level sort

Grouping

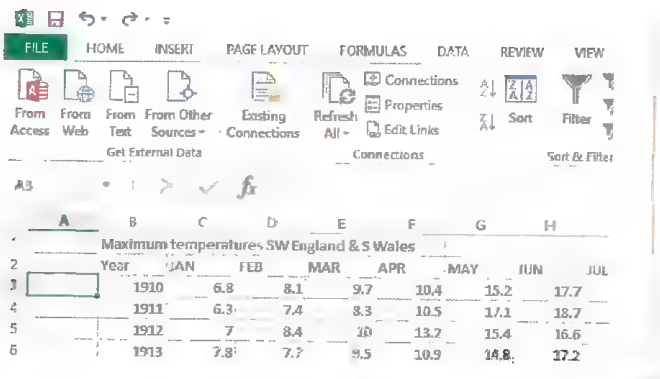
Grouping is an Excel® feature, similar to outlining in a word processor, making it easy to show or hide groups (sections) of your data. At the bottom of each group, there will be a section where functions show useful information such as the subtotal, average or count.

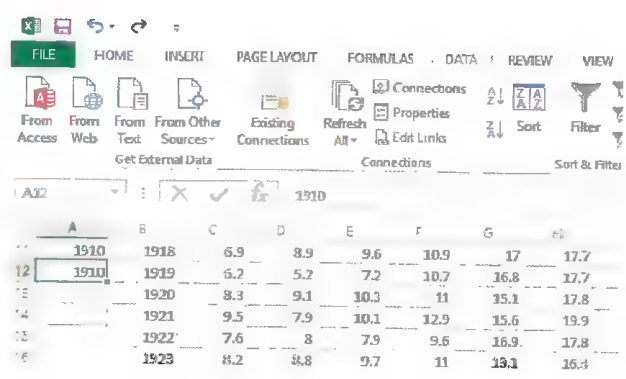
Step by step: Grouping

1 This step by step shows how to enter and explore the set of data shown in the figure below, which can be downloaded from the UK Met Office website.

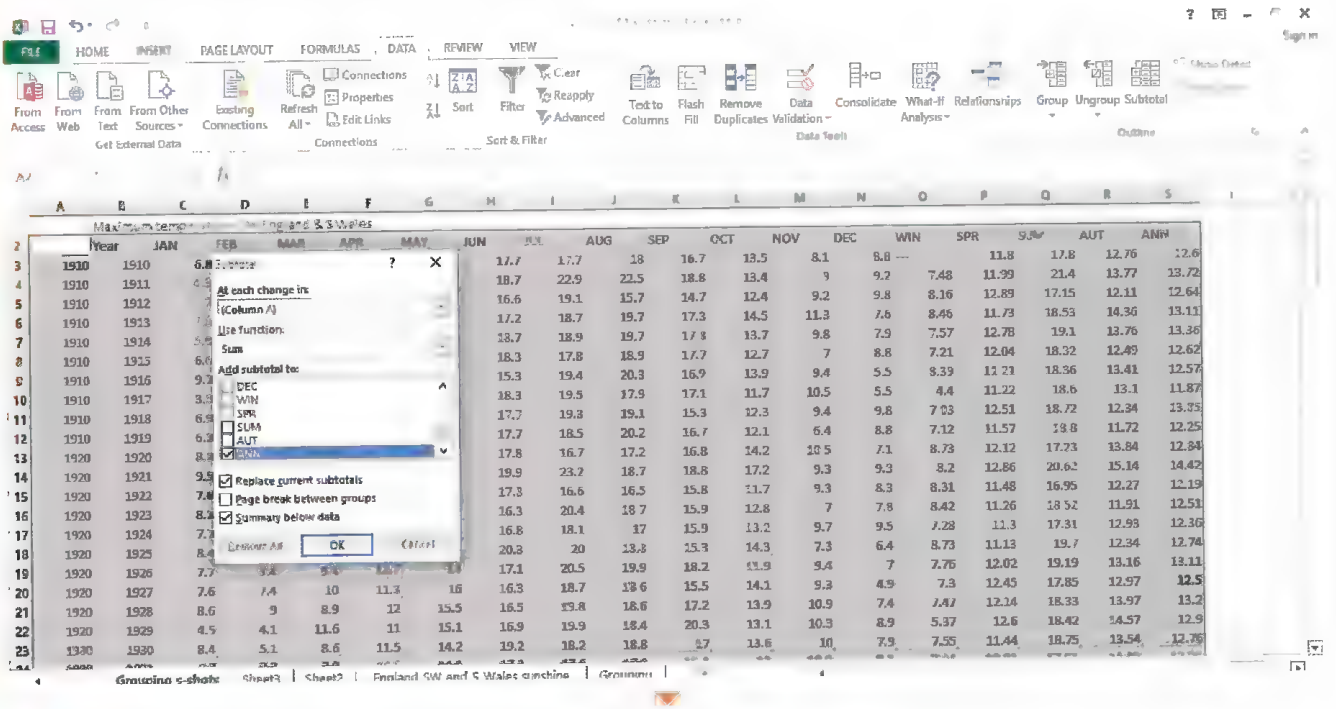


2 You need to insert a new column that holds data that can be used for grouping. We will use this column for another set of years so we can group by decade. Insert the new column to the left by placing the cursor anywhere in the current column A, then either right click, select insert, select entire column or use the keyboard by pressing <Alt> <I> then <C>. Type '1910' (the first decade), into A3, then copy down to A12 (all the years to be in this group). Repeat this for the other decades, so that each group has the start year for that decade in the A column.

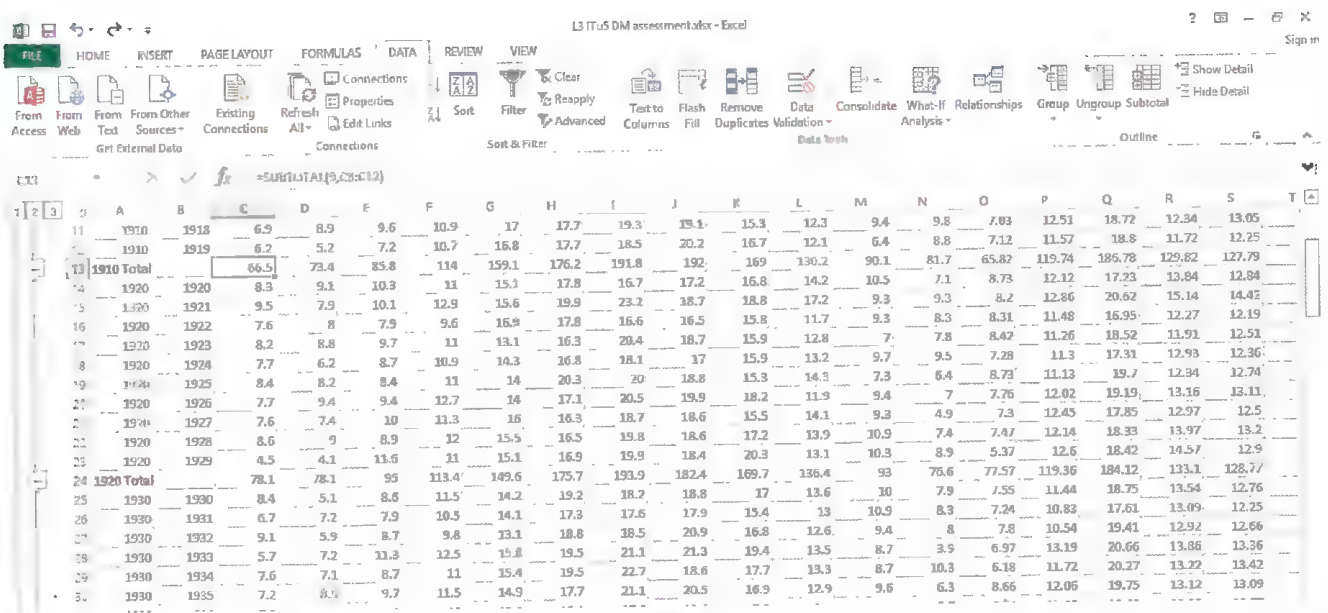




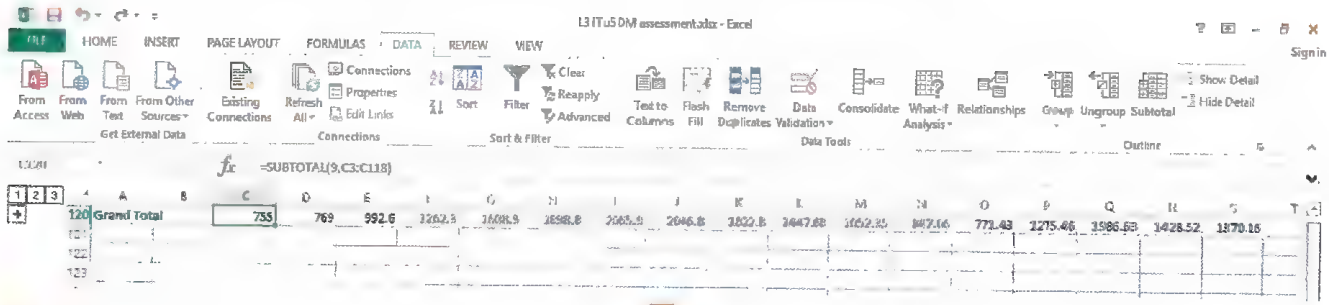
4 Place the cursor anywhere in the data, eg B3, then click on the Subtotal button in the Data toolbar. Excel® will highlight the data it finds around B3 and show the Subtotal dialog box. You can choose the function used for each group, (Sum, the default is shown in this screenshot). If you want more statistics, you could insert rows between the subtotals and groups later. Use Add subtotal to check boxes in this dialog box to select all columns where you want the function. Click the OK button to complete.



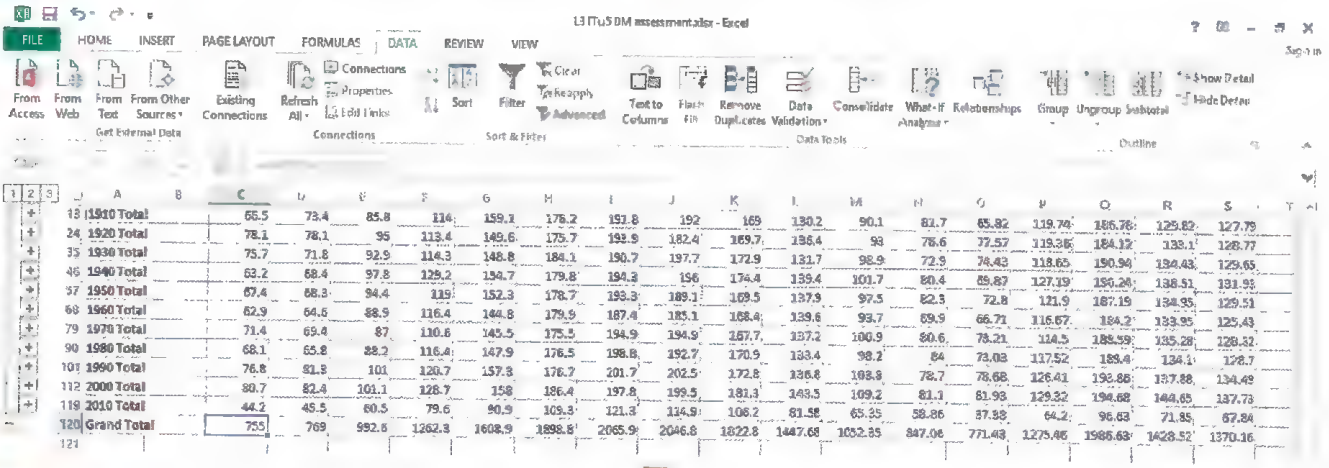
5 You now have subtotals and Excel® has also grouped your data. In this example, the =SUBTOTAL() function has 9 as the first argument because we selected SUM from the Add subtotal to dialog box in step 3. The grouping is shown to the left of the row numbers with the numbers (1, 2, 3) used to select how much of the grouping you can see on the spreadsheet.



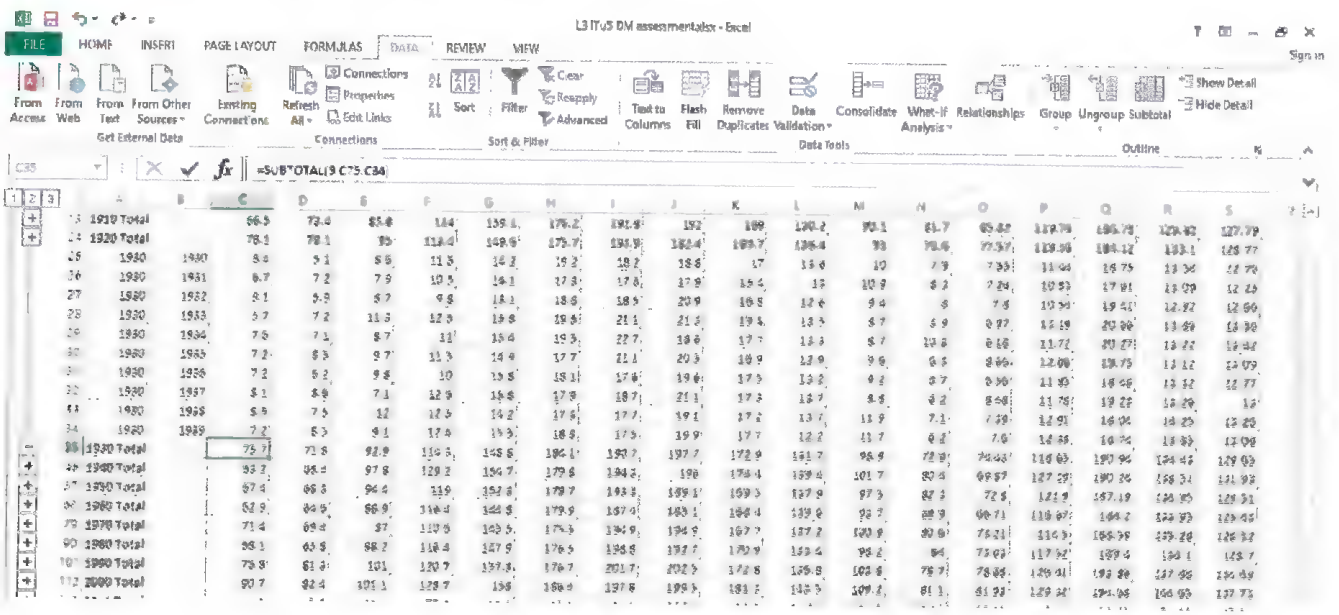
5 Click on the 1 in grouping to see just the summary row for all the whole group.



Click on the + under 1 in grouping to see the summary rows for all the groups.



Click on a + under 2 in grouping to see the detail of that group.



Filtering

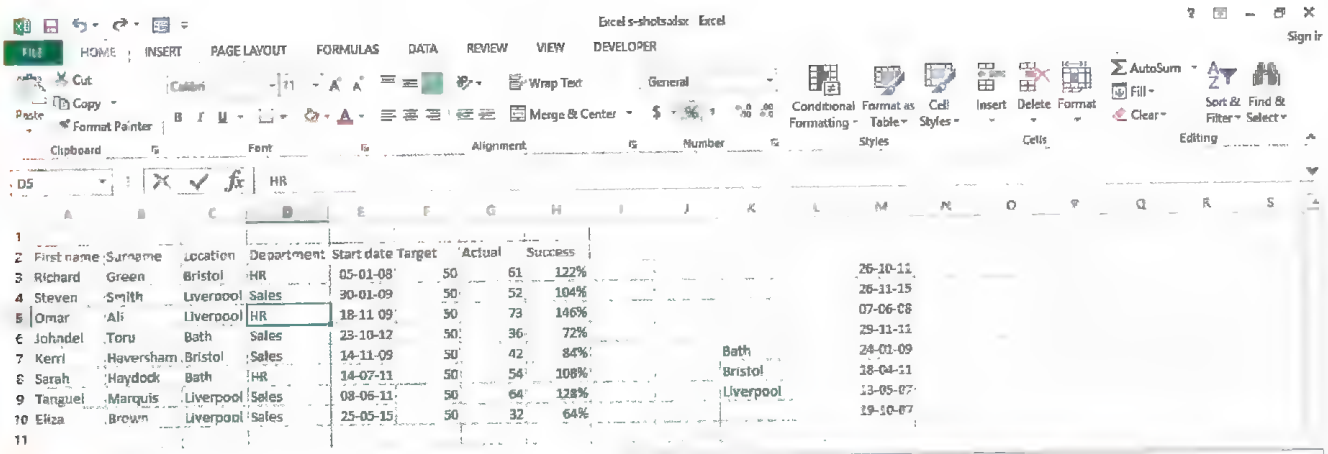
Filtering allows column headings on a set of data to reduce (filter) the rows shown in the spreadsheet.

Pivot tables

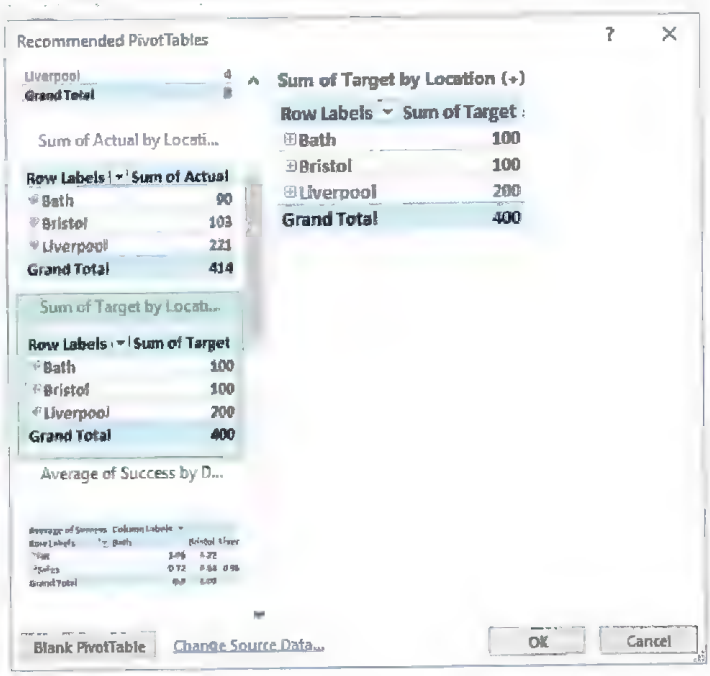
Excel® pivot tables help you to explore large spreadsheet data sets by copying the parts you want to a sheet where you can choose row and column headings for this summary and apply further filtering if needed.

Step by step: Pivoting data

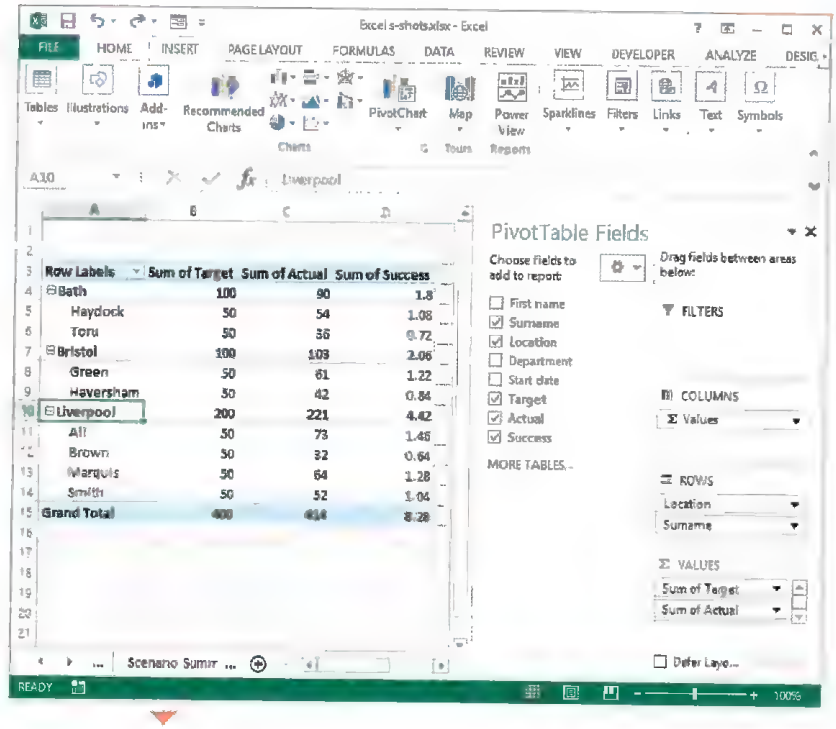
1 This step by step shows how to insert a pivot table into a spreadsheet and how to create a pivot chart. You need some data in your spreadsheet for the pivot chart to use. The data should have column headings (no need for any formatting for these) and be a mix of text and numbers.



2 Select any cell in your data then use the Recommended PivotTables button in the Insert toolbar. This has two major advantages over using the PivotTable button: it checks that your data is suitable for use as a pivot table and it gives you some previews of potential layouts for your pivot table. Select the layout closest to what you want, then use the OK button.



A new worksheet is created for the pivot chart you chose.
You can rearrange the pivot table by dragging fields in the PivotTable Fields control to the right of the pivot table display. The example here needs formatting in the Sum of Success column to show the numbers as percentages. To do this, right click on the column, select Value Field Settings, then select Number Format, where you can set the format of this column to percentage. The Grand Total and sub-totals of this column would be more meaningful as averages.
To do this, right click on the column, select Summarize Values By, and then select Average.



You can create a chart from your pivot table by simply clicking on the Pivot Chart button in the Insert toolbar. This chart is dynamic, with buttons that you can use to change the view of the data charted. The example shown here allows you choose which Locations are shown.

Importing data

When data is brought in from another source, it is said to be imported into a spreadsheet, usually as a file that the source app saved in a format that Excel® can recognise. Imports are usually **text files** as these do not include any extra formatting information that would not be recognised by the spreadsheet. Each line of the text file will be a row in the spreadsheet, so there needs to be a method used by the spreadsheet to break the line into cells. There are two methods used to define cells in a row: fixed width and delimiter characters.

- Fixed width is very simple, for example first 8 characters into the first cell (column), 3 characters into the next, and so on.
- Delimiter characters are used in the text file between items of data to indicate how to allocate them to cells. Tab delimited files use <Tab> to separate the data in each line and are usually given a **TSV** extension when saved. **CSV** is the other commonly used import data file type, with commas used to delimit items of data.

Key terms

Text file – a text file is a simple document with no formatting. They are just text, unlike apps such as Excel® or Word® which create documents that have lots of extra information (formatting) such as page size, fonts, images and tables which makes the document more complex and can be opened by an app that understands that format.

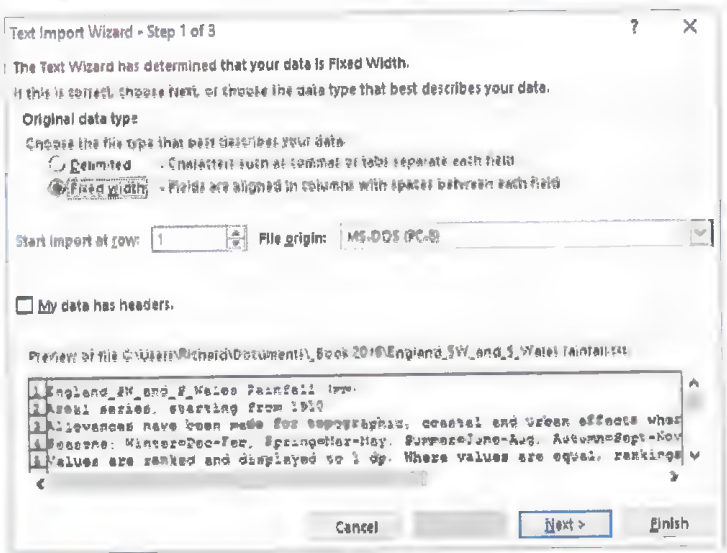
CSV file – a simple text file that can be used to transfer data between apps. A comma is used between each item on a line to separate them, so if they were imported into a spreadsheet they would go into different cells. Hence CSV stands for 'comma separated values'.

TSV file – a simple text file that can be used to transfer data between apps. A <Tab> character is used between each item on a line to separate them, so if they were imported into a spreadsheet they would go into different cells. Hence TSV stands for 'tab separated values'.

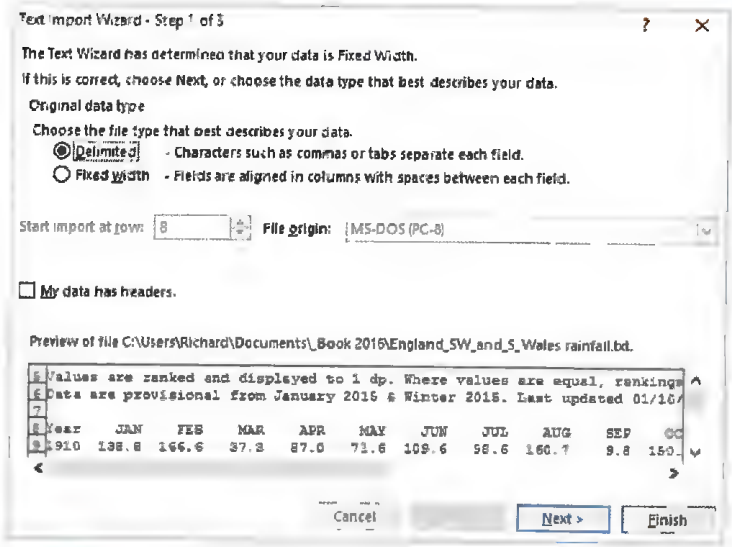
Step by step: Importing data

1 This step by step shows how to import a text file that has been downloaded from the Met Office website into an Excel® spreadsheet. Select the cell (A3) where you want the top left corner of the data import. Click on the From Text button in the Data toolbar to bring up a dialog box where you can select your data file.

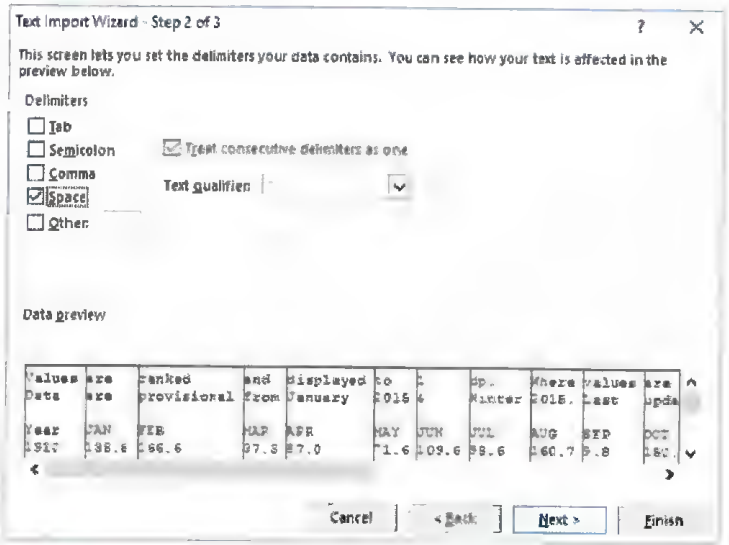
2 After you have double clicked on the text data file, the wizard starts. You see a preview of the text file. The text data file shown here starts with some lines of text describing the data, and these do not need to be imported into Excel®.



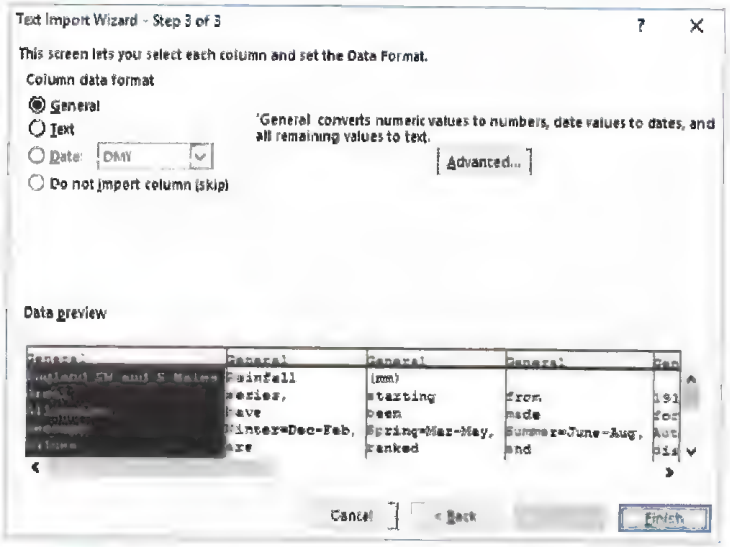
3 Scrolling down the text data preview shows that the data actually starts on line 8, so the Start import at row 1 needs to be changed to row 8. Select Delimited (as there are spaces between data items in the text file), then use the Next button.



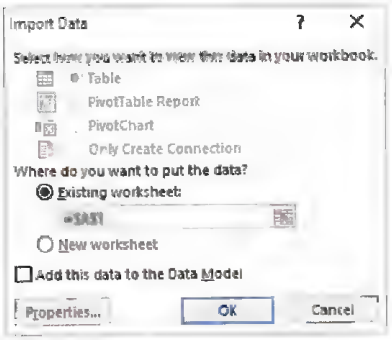
4 The Delimiters need to be set to Space and the Treat consecutive delimiters as one need to be selected as there are variable numbers of spaces between the data items. The preview shows where the data will be divided into cells when imported into the spreadsheet.



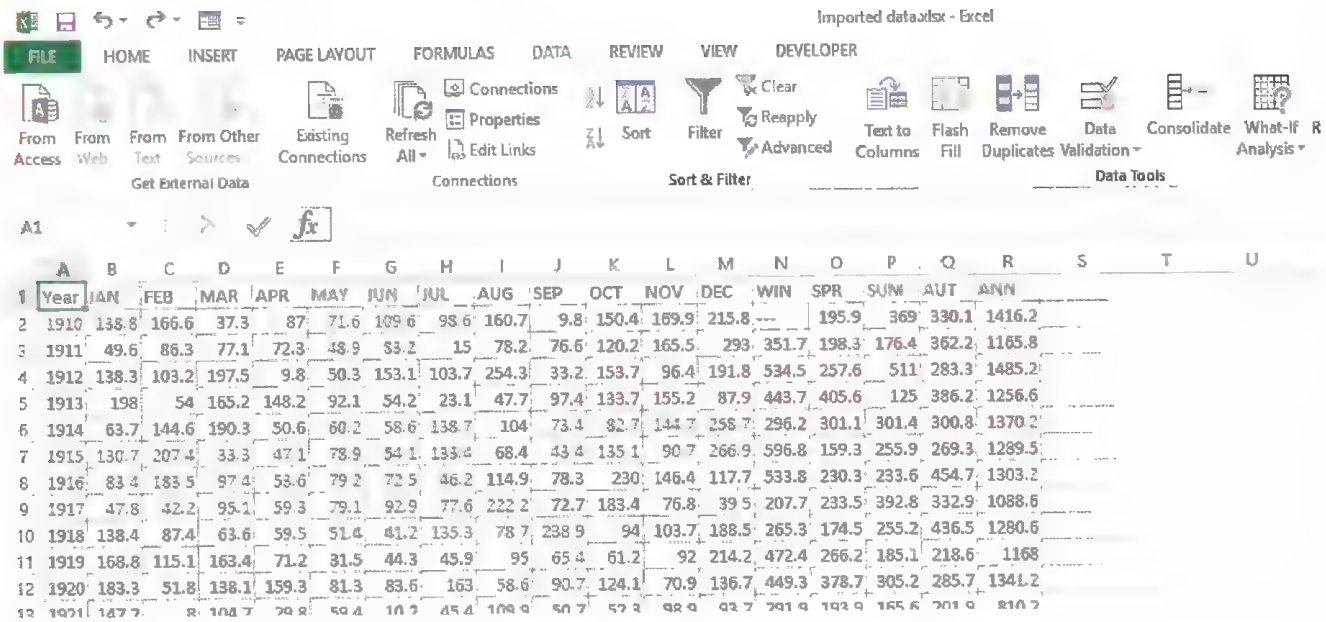
5 You could fine tune the formatting that Excel® will apply to the cells when the data file is imported at step 3 of the wizard. You can leave the default settings and then click Finish.



6 Confirm cell (A3) is where you want the data import to start from and then click OK.



7 The data file has been imported. You should move the cursor around some of the cells to check that the text data items have been allocated to individual cells.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	WIN	SPR	SUM	AUT	ANN			
2	1910	138.8	166.6	37.3	87	71.6	109.6	98.6	160.7	9.8	150.4	169.9	215.8	---	195.9	369	330.1	1416.2			
3	1911	49.6	86.3	77.1	72.3	48.9	53.2	15	78.2	76.6	120.2	165.5	293	351.7	198.3	176.4	362.2	1165.8			
4	1912	138.3	103.2	197.5	9.8	50.3	153.1	103.7	254.3	33.2	153.7	96.4	191.8	534.5	257.6	511	283.3	1485.2			
5	1913	198	54	165.2	148.2	92.1	54.2	23.1	47.7	97.4	133.7	155.2	87.9	443.7	405.6	125	386.2	1256.6			
6	1914	63.7	144.6	190.3	50.6	60.2	58.6	138.7	104	73.4	82.7	144.7	258.7	296.2	301.1	301.4	300.8	1370.2			
7	1915	130.7	207.4	33.3	47.1	78.9	54.1	133.4	68.4	43.4	135.1	90.7	266.9	596.8	159.3	255.9	269.3	1289.5			
8	1916	83.4	183.5	97.4	53.6	79.2	72.5	46.2	114.9	78.3	230	146.4	117.7	533.8	230.3	233.6	454.7	1303.2			
9	1917	47.8	42.2	95.1	59.3	79.1	92.9	77.6	222.2	72.7	183.4	76.8	39.5	207.7	233.5	392.8	332.9	1088.6			
10	1918	138.4	87.4	63.6	59.5	51.4	41.2	135.3	78.7	238.9	94	103.7	188.5	265.3	174.5	255.2	436.5	1280.6			
11	1919	168.8	115.1	163.4	71.2	31.5	44.3	45.9	95	65.4	61.2	92	214.2	472.4	266.2	185.1	218.6	1168			
12	1920	183.3	51.8	138.1	159.3	81.3	83.6	163	58.6	90.7	124.1	70.9	136.7	449.3	378.7	305.2	285.7	1341.2			
13	1921	147.2		8	104.7	29.8	59.4	10.2	45.4	109.9	50.7	52.3	98.9	93.7	291.9	193.9	165.6	201.9	810.2		

Exporting data

Exporting data involves saving a worksheet in Excel® in a format that can be understood by a different app and importing it into that app. As with importing data, simple text files are usually used for exports, usually with a delimiter character to separate the cells as they get saved.

To export from Excel®, simply use the File, Save as dialog box when the type of file can be chosen. Excel® will save comma separated values with a CSV extension, whereas tab separated values are saved with a default TXT extension.

Autofill

The autofill feature of Excel® is there to automatically enter data into cells for you. You just need to type enough into Excel® to start, select it and then use the mouse to drag the bottom right handle (corner) of the selection to where you want the autofill (see Table 5.5 Excel® autofill).

Table 5.5: Excel® autofill

Initial selection	Extended series
1, 2, 3	4, 5, 6,..
9:00	10:00, 11:00, 12:00,..
Mon	Tue, Wed, Thu,..
Monday	Tuesday, Wednesday, Thursday,..
Jan	Feb, Mar, Apr,..
Jan, Apr	Jul, Oct, Jan,..
Jan-07, Apr-07	Jul-07, Oct-07, Jan-08,..
15-Jan, 15-Apr	15-Jul, 15-Oct,..
2007, 2008	2009, 2010, 2011,..
1-Jan, 1-Mar	1-May, 1-Jul, 1-Sep,..
Qtr3 (or Q3 or Quarter3)	Qtr4, Qtr1, Qtr2,..

Replication

Formulae and functions often need to be replicated into new rows or columns of data. In most situations, this will be using copy and paste or dragging the bottom right handle of the highlighted cell(s) containing the calculation(s). Replication is often needed after data is imported to recreate any calculations that were lost in the text file.

Using macros and buttons to initiate procedures

Macros are the program code that can work with your spreadsheet to automate actions. Excel® provides a rich programming environment called Visual Basic® for Applications (VBA) which can match many dedicated programming languages for capability. You can use buttons or other objects such as images to initiate macro code procedures as well as allocating key press combinations to start your programs.

Research

Using macros is a massive topic that deserves further research, reading and practice.

Many VBA programmers use Excel® to record a macro where they carry out a spreadsheet operation they want to automate. The code produced by this recording can then be edited into exactly what the programmer wants and can be copied to another part of the VBA program code where it is needed.

To use macros, you need the Developer toolbar which Excel® does not show by default. This toolbar can be shown by using the File, Options menu to bring up the Excel® Options dialog where you select the Customize Ribbon option. The Developer checkbox under Customize the Ribbon options has to be ticked and then you use the OK button to exit Excel® Options to return to the workbook where the Developer toolbar will now be shown.

The Developer toolbar has a button with which you can start recording your actions as a macro. This changes into a button to stop recording when you have completed your macro. It is a good idea to practise your actions before recording a macro to reduce mistakes. The actions that you record can be just about anything that can be done to the spreadsheet with the mouse or keyboard.

The Developer toolbar has a button named Visual Basic which can take you to the VBA macro code environment. This environment will run as a separate app to your spreadsheet and will end when the spreadsheet is closed. Many programmers use the VBA macro code environment to edit their macros to delete mistakes in a recording or to add more functionality to their macro programs.

Data entry and validation

Data entry and validation requirements and methods include:

- using data entry forms
- restricting data input to acceptable values
- protecting cells by hiding, locking and password protecting
- ease of use techniques
- automated data transfer between sheets or applications
- adding user prompts and messages.

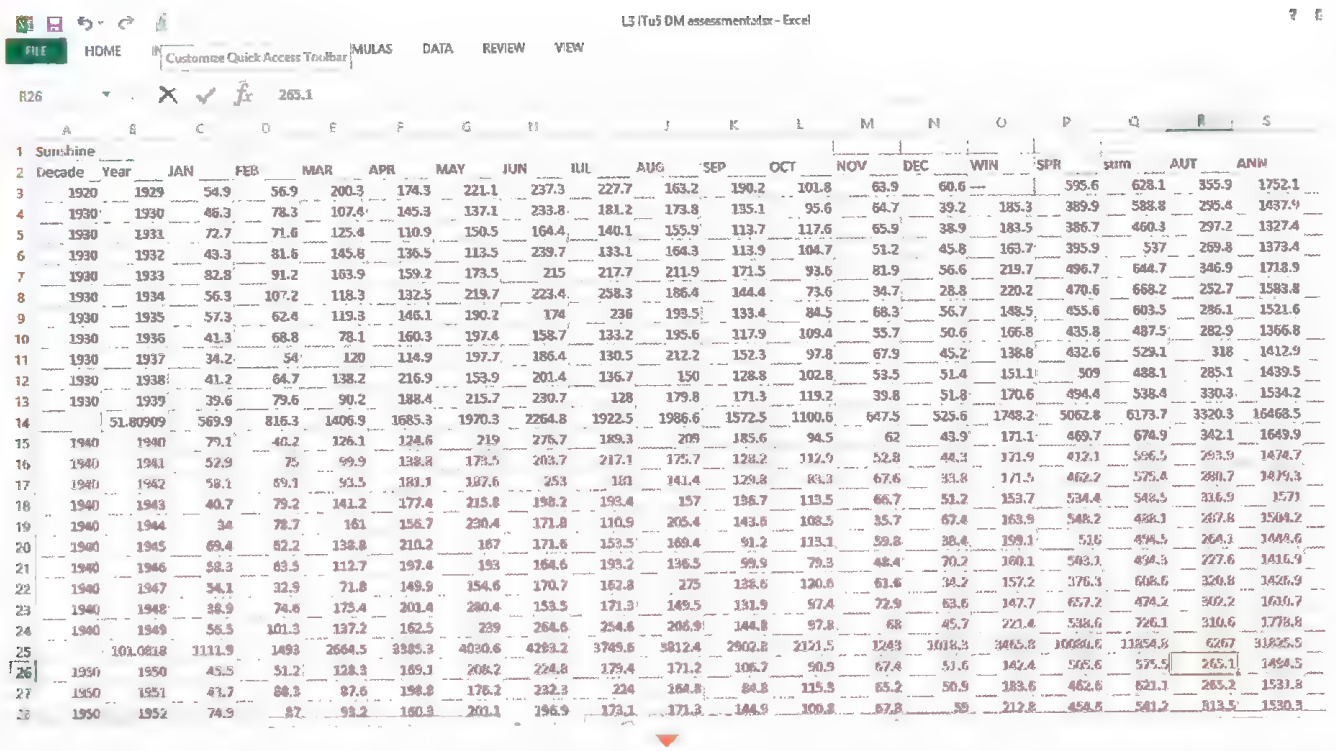
Data entry forms

Many users prefer to enter or edit a large spreadsheet with a data entry form to show all the entries in one form, rather than needing to scroll around the rows and columns. A data entry form is quickly and easily produced by Excel®. All you need to do is select a cell anywhere inside some data with column headings, then click on the Form button. Recent versions of Excel® do not include the Form button in the toolbars by default, so you will need to customise a toolbar yourself (see the Step by step).

A data entry form can be used to find a particular record for editing or deleting by using the Criteria button on the form. Searching can use the * and ? wildcards to help find the required records. A search for Pet* would bring up everything starting with Pet, for example Pete or Petra, because the * wildcard represents any number of characters. A search for Pet? brings up everything starting with Pet which is four characters long, for example Pete but not Petra, because the ? wildcard represents one character.

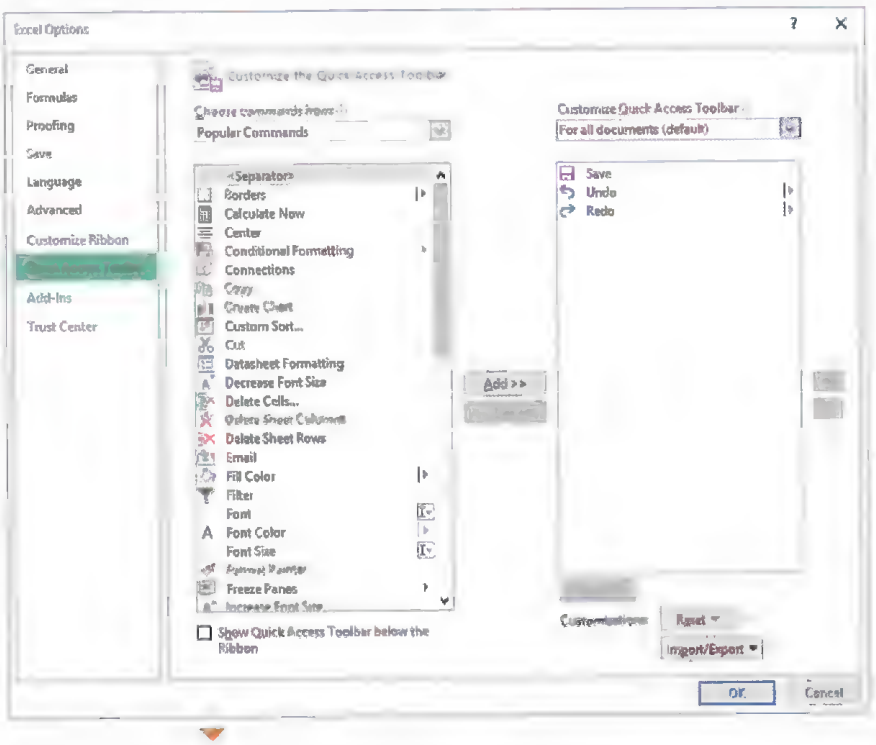
Step by step: Creating a data entry form

1 This step by step shows how to add a Form button to the Excel® toolbars and use this to bring up a data entry form for some worksheet data. The data entry form is then used to find a record in the data. Click on Customize Quick Access Toolbar.



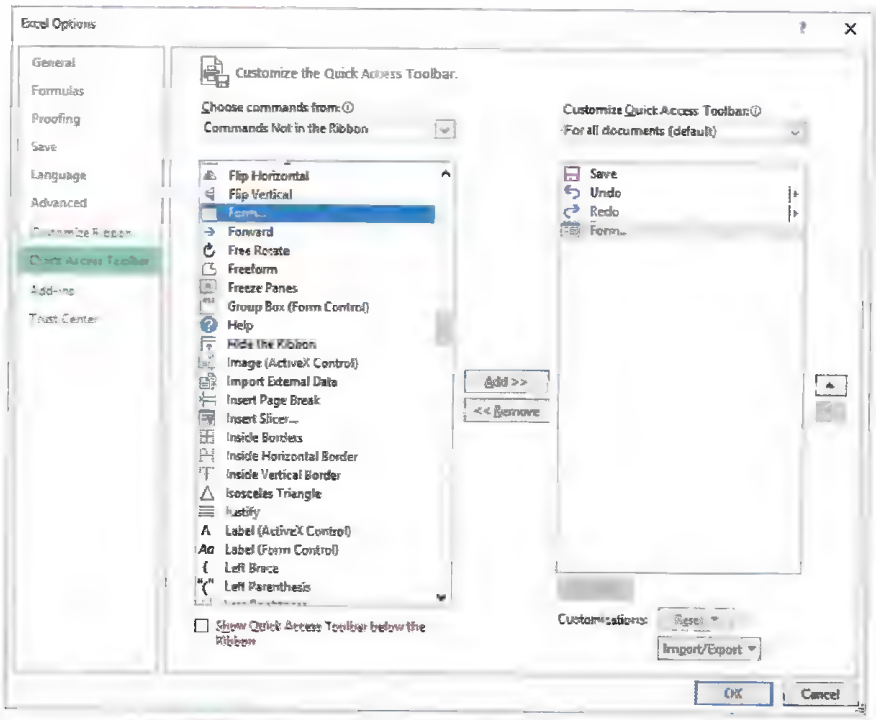
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Sunshine																		
2	Decade	Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	WIN	SPR	sum	AUT	ANN
3	1920	1929	54.9	56.9	200.3	174.3	221.1	237.3	227.7	163.2	190.2	101.8	63.9	60.6		595.6	628.1	355.9	1752.1
4	1930	1930	46.3	78.3	107.4	145.3	137.1	233.8	181.2	173.8	135.1	95.6	64.7	39.2	185.3	389.9	588.8	295.6	1437.9
5	1930	1931	72.7	71.6	125.4	110.9	150.5	164.4	140.1	155.9	113.7	117.6	65.9	38.9	183.5	386.7	460.3	297.2	1327.4
6	1930	1932	43.3	81.6	145.8	136.5	113.5	239.7	133.1	164.3	113.9	104.7	51.2	45.8	163.7	395.9	537	269.8	1373.4
7	1930	1933	82.8	91.2	163.9	159.2	173.5	215	217.7	211.9	171.5	93.6	81.9	56.6	219.7	496.7	644.7	346.9	1718.9
8	1930	1934	56.3	107.2	118.3	132.5	219.7	223.4	258.3	186.4	144.4	73.6	34.7	28.8	220.2	470.6	668.2	252.7	1583.8
9	1930	1935	57.3	62.6	119.3	146.1	190.2	174	236	193.5	133.4	84.5	68.3	56.7	148.5	455.6	603.5	286.1	1521.6
10	1930	1936	41.3	68.8	78.1	160.3	197.4	158.7	133.2	195.6	117.9	109.4	55.7	50.6	166.8	435.8	487.5	282.9	1366.8
11	1930	1937	34.2	54	120	114.9	197.7	186.4	130.5	212.2	152.3	97.8	67.9	45.2	138.8	432.6	529.1	318	1412.9
12	1930	1938	41.2	64.7	138.2	216.9	153.9	201.4	136.7	150	128.8	102.8	53.5	51.4	151.1	509	488.1	285.1	1439.5
13	1930	1939	39.6	79.6	90.2	188.4	215.7	230.7	128	179.8	171.3	119.2	39.8	51.8	170.6	494.4	538.4	330.3	1534.2
14		51.80909	569.9	816.3	1406.9	1685.3	1970.3	2264.8	1922.5	1986.6	1572.5	1100.6	647.5	525.6	1748.2	5062.8	6173.7	3320.3	16468.5
15	1940	1940	79.1	40.2	126.1	124.6	219	275.7	189.3	209	185.6	94.5	62	43.9	171.1	469.7	674.9	342.1	1649.9
16	1940	1941	52.9	75	99.9	138.8	173.5	203.7	217.1	175.7	128.2	112.9	52.8	44.3	171.9	412.1	596.5	293.9	1474.7
17	1940	1942	58.1	69.1	93.5	181.1	187.6	253	181	141.4	129.8	83.3	67.6	33.8	171.5	462.7	575.4	280.7	1479.3
18	1940	1943	40.7	79.2	141.2	177.4	215.8	198.2	193.4	157	136.7	113.5	66.7	51.2	153.7	534.4	548.5	316.9	1573
19	1940	1944	34	78.7	161	156.7	230.4	171.8	110.9	205.4	143.8	108.5	35.7	67.4	163.9	548.2	488.1	287.8	1504.2
20	1940	1945	69.4	62.2	138.8	210.2	167	171.6	153.5	169.4	91.2	113.1	59.8	38.4	199.1	516	494.5	264.3	1448.6
21	1940	1946	58.3	63.5	112.7	197.4	193	164.6	193.2	136.5	99.9	79.3	48.4	70.2	160.1	503.1	494.5	227.6	1416.9
22	1940	1947	54.1	32.9	71.8	149.9	154.6	170.7	162.8	275	138.6	120.6	61.6	34.2	157.2	376.3	608.6	320.8	1426.9
23	1940	1948	38.9	74.6	175.4	201.4	280.4	153.5	171.3	149.5	131.9	97.4	72.9	63.6	147.7	657.2	474.2	302.2	1630.7
24	1940	1949	56.5	101.3	137.2	162.5	239	264.6	254.6	206.9	144.8	97.8	68	45.7	221.4	538.6	726.1	310.6	1778.8
25		101.0818	1111.9	1493	2664.5	3385.3	4030.6	4263.2	3749.6	3812.4	2902.8	2121.5	1243	1018.3	3465.8	10080.6	11854.6	6267	31825.5
26	1950	1950	45.5	51.2	128.3	169.1	208.2	224.8	179.4	171.2	106.7	90.9	67.4	51.6	142.4	506.6	575.5	265.1	1454.5
27	1950	1951	41.7	88.3	87.6	198.8	176.2	232.3	224	184.8	84.8	115.9	65.2	50.9	183.6	462.6	621.1	265.2	1531.8
28	1950	1952	74.9	87	93.2	160.9	201.1	196.9	173.1	171.3	144.9	100.8	67.8	55	212.8	458.5	581.2	313.5	1530.3

2 Click on More commands from the Customize Quick Access Toolbar menu to bring up the Excel® options dialog box showing the Quick Access Toolbar options.

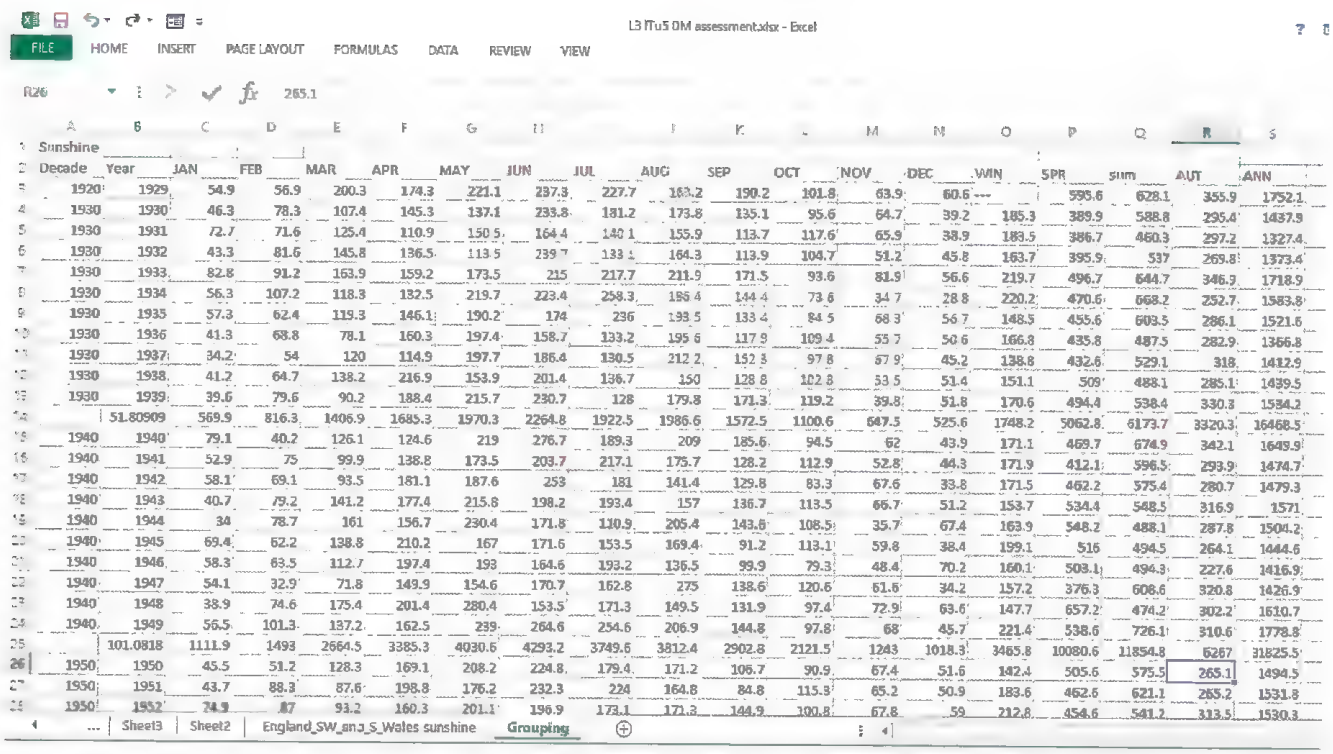


3 Select Commands Not in the Ribbon from the Choose commands from drop-down list box. This shows the buttons that you can add to the toolbar. Scroll down these so that the Form button is visible.

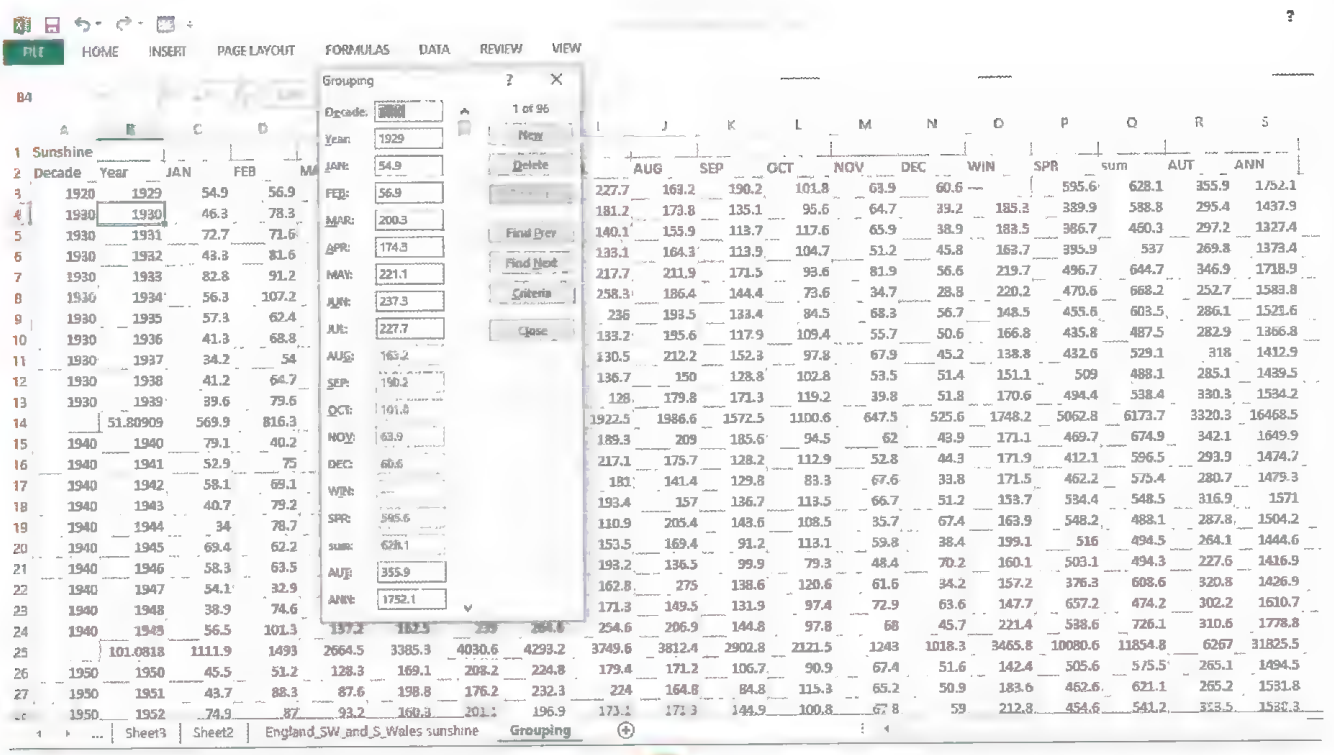
4 Click on the Form button then use the Add button to include this in the Quick Access Toolbar.



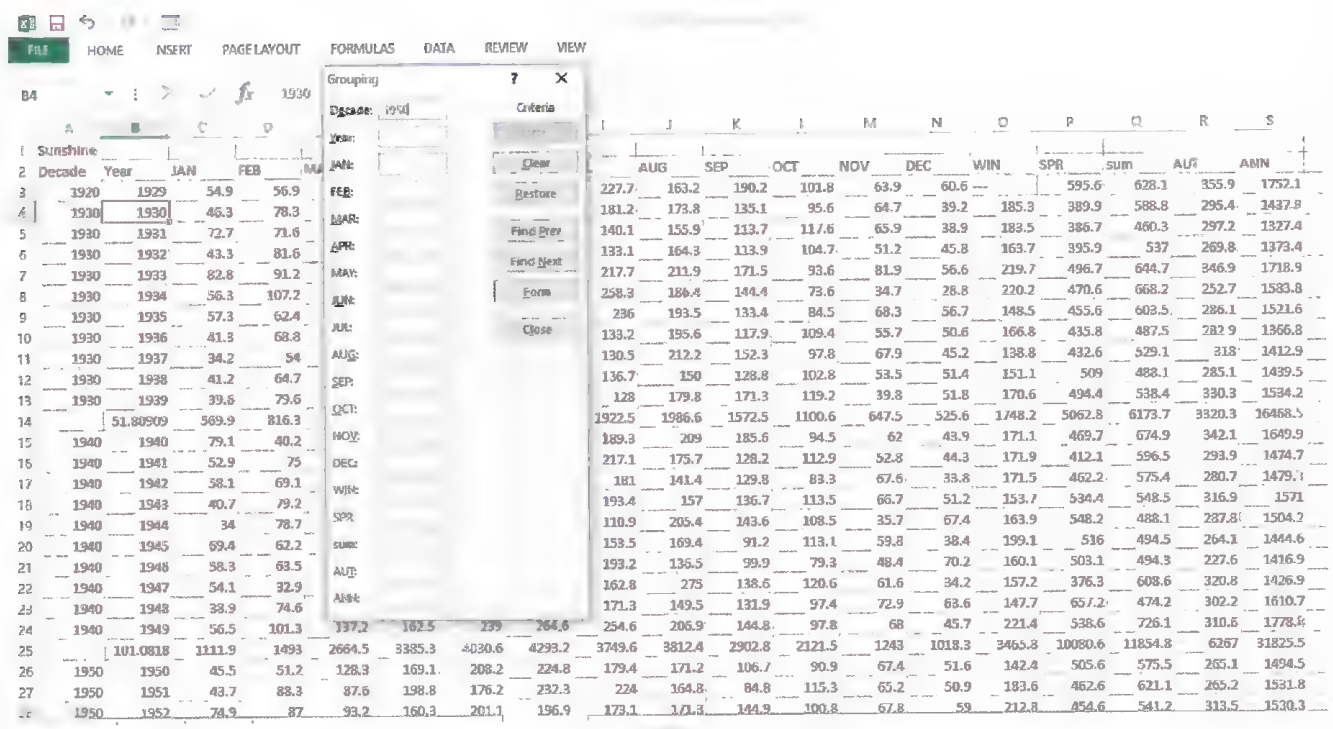
5 Click on the OK button. You now have the Form button in the Quick Access Toolbar.



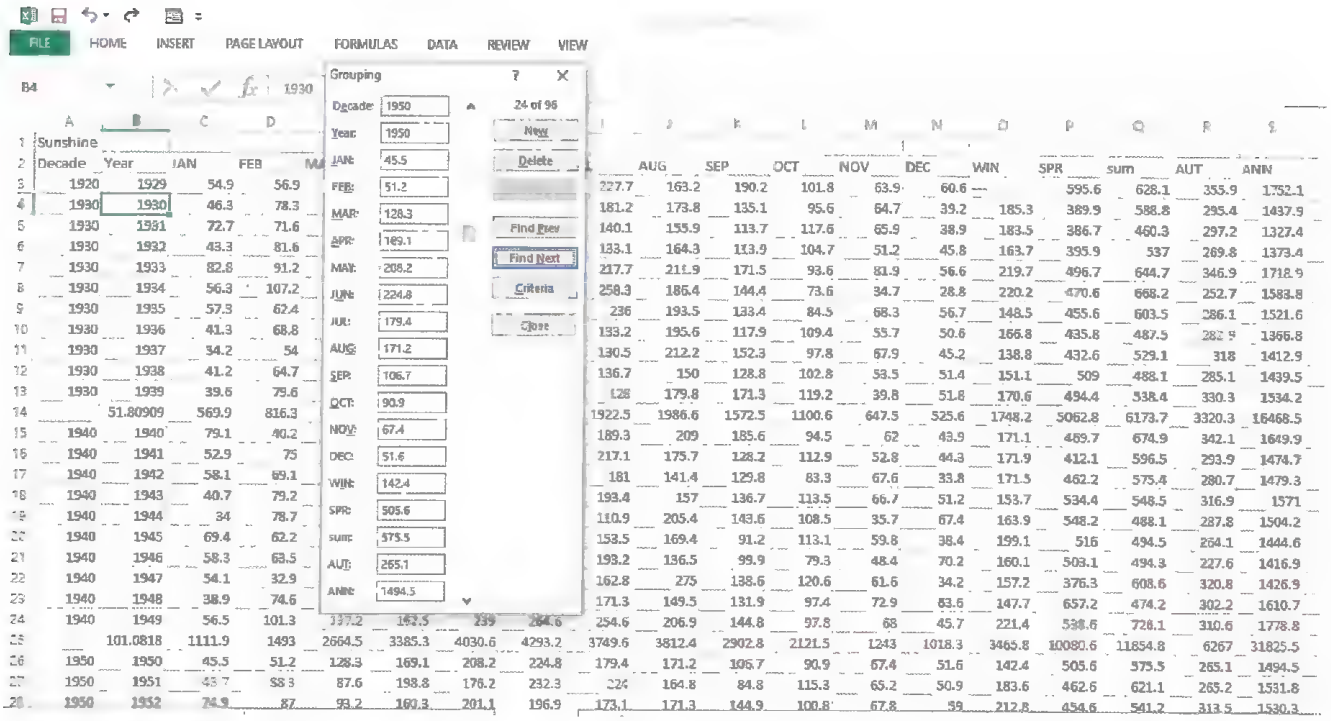
Now that you have the Form button available, you can show the data entry form. To do this, select any cell inside the data in the worksheet then click on the Form button.



You can use the Criteria button to find a record row in the data. To do this, click on the Criteria button; the form clears. Type '1950' into the Decade text box.



4 Click on the Find Next button. The data entry form shows the first record which has 1950 in decade. Notice that the cursor does not move in the worksheet behind the form. You could edit any of the cells shown in the data entry form. The Find Next button can be used to show the next record in the data entry form, or the Find Prev to show the previous record. Use the Close button when you are finished with the data entry form.

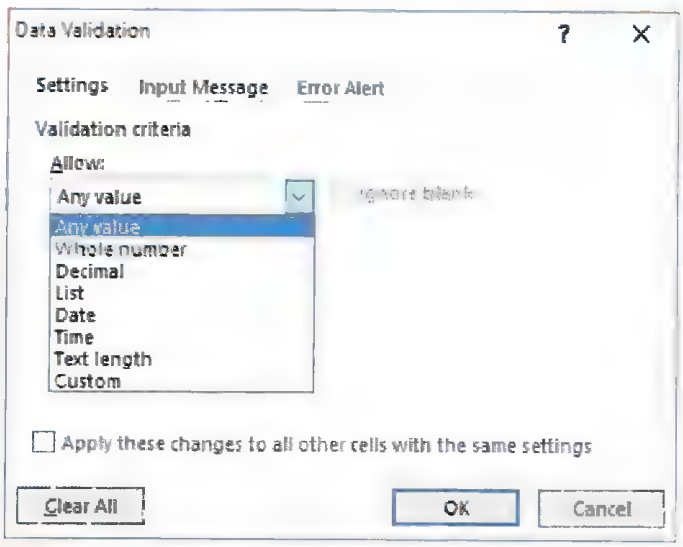


Restricting data input

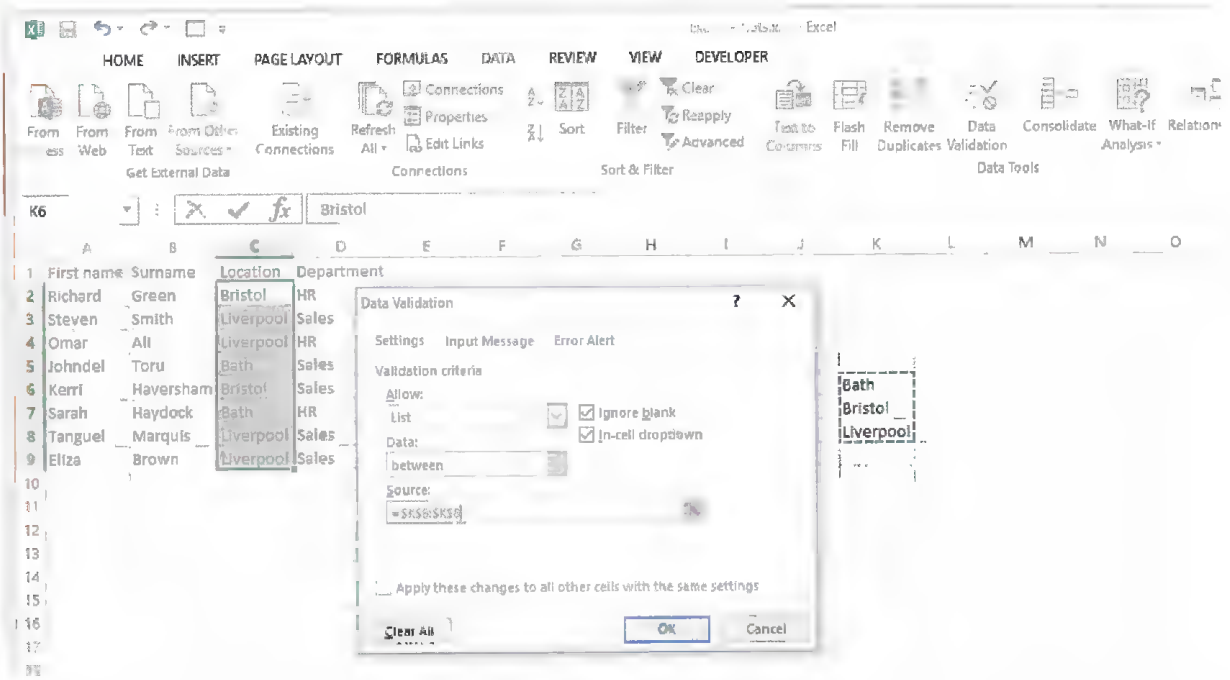
Most errors in a data set are from mistyping. Excel® can help reduce these typos by restricting data input into selected cells to acceptable values.

Step by Step: Restricting data input to acceptable values

1 To do this, first select the cells where you want to set validation on data entries then use the use the Data Validation drop-down menu on the Data toolbar to select Data Validation.
In the dialog box you can choose what to allow to be entered into the selected cells.



The Data Validation dialog box gives good control over messages to the user on what data can be entered and to alert them to any mistyping with a message on what to do about it. You can restrict data entry into the cells to a list. The result of this example is to let the user choose from a drop-down list of values from cells K6 to K8 when clicking into cells C2 to C9.

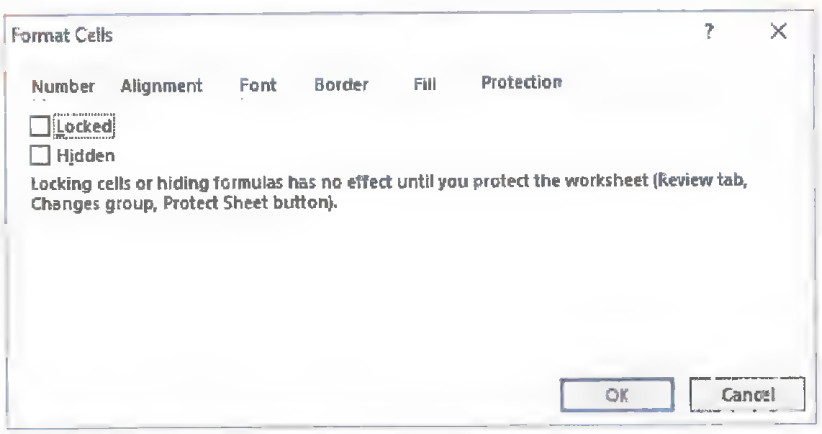


Protecting cells

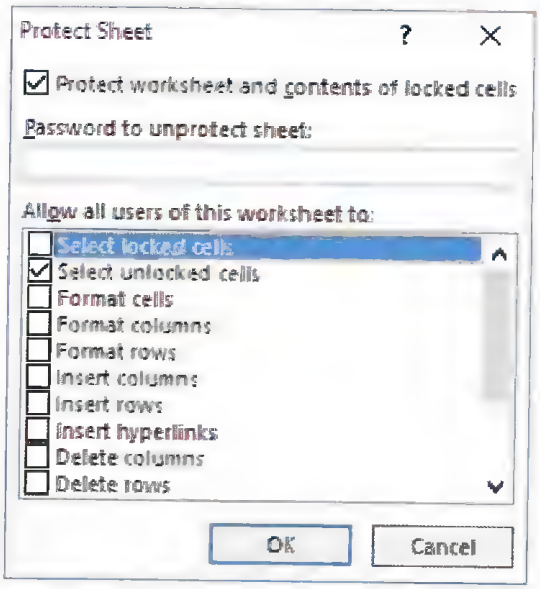
Many spreadsheets holding data and calculations use cell protection to prevent users accidentally typing over calculations and to help data entry. Protecting cells can assist data entry by only allowing the user to select cells which are unprotected. This can make the spreadsheet much quicker to use as the <Arrow> keys jump to the next cell where data is to be input.

Step by step: Protecting cells

To use protection, you need to format all the cells which can be edited then use the Review toolbar to protect the worksheet. Cells available for editing are selected; a right click brings up the Format Cells dialog box then you need to select the Protection tab. You can then choose to select Locked to lock the cells or, as shown here, deselect the Locked checkbox to unlock cells.



When all the cells needed for editing have been formatted as not locked, you can use the Review toolbar to click on Protect Sheet to bring up the Protect Sheet dialog box. Here you can untick the Select locked cells checkbox so that the user is restricted to the cells that have been formatted as unlocked. You can also set a password using this dialog box. (To remove the protection, use the same button on the Review toolbar, which will now show Unprotect sheet.) Cells can also be protected by hiding the row or column so that they cannot be seen or easily selected.



Ease of use techniques

Excel® offers you a good choice of techniques to make a spreadsheet easier to use. These include drop-down menus and list boxes. A drop-down menu lets you choose from a list of choices when entering data into a cell. This makes data entry faster and errors less likely as you would not be able to mistype.

Before you can set up a drop-down menu, you need to have the menu choices entered into a range of cells with no gaps. Select the cells to have the drop-down menu, then use the Data Validation in the Data toolbar to bring up the dialog box (see step 1 of Step by Step: Restricting data input to acceptable values). Click on List in the Allow drop-down list of this dialog box to show Figure 5.9 where you can select the source cells holding the menu choices (=$B$3:$B$6 in this example). Click on the OK button to complete.

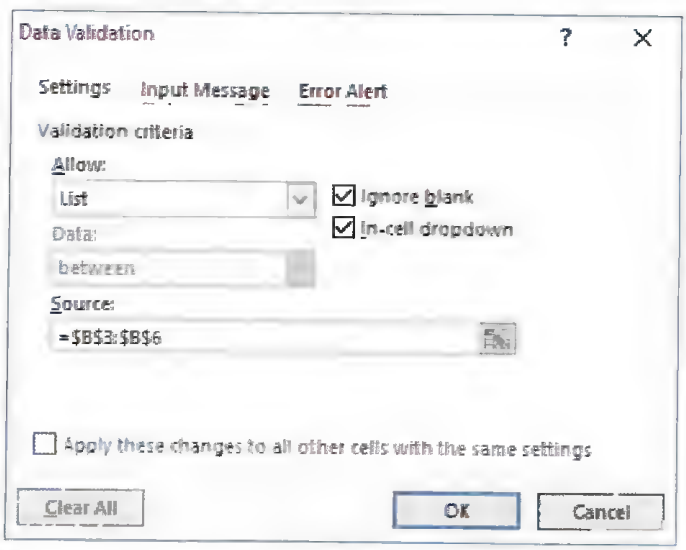


Figure 5.9: Setting up a drop-down menu

Automated data transfers

There are many methods used to automate data transfer between sheets and applications.

Simple formula

If all you need is to show and use data from one worksheet with another worksheet in the same spreadsheet (workbook), you can use a simple formula. To do this, select the cell where you want the other worksheet data, type = then select the cell in the other worksheet with the data. Use <Enter> to complete. You can easily copy and paste this into other cells if needed. This simple formula will also work between spreadsheets. If you do this, be careful not to move or rename the spreadsheets involved. When you open the workbook spreadsheet containing these formulae, you will be prompted to update the links to the original (source) workbook (see Figure 5.10).

Figure 5.10: Update links prompt

Linking

A common method for automating between apps is to use linking. A section of an Excel® spreadsheet can be copied and then **Paste Special** can be used to paste this into Word® as a link (see Figure 5.11) There are several choices of how you want the linked spreadsheet to appear in the Word document. Each time the document is opened, you will be prompted on whether to update the link to show current information in the document.

A macro can be used to automate data transfers. As the macro is program code, you have several options for tweaking the data transfer, for example by adding calculations, formatting cells, or by deleting rows or columns.

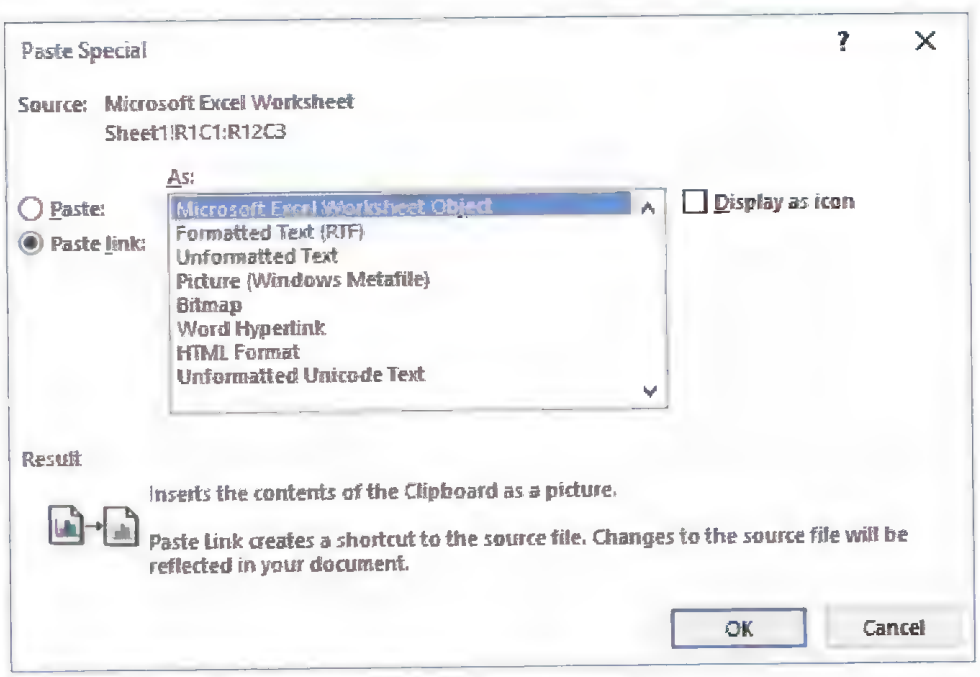


Figure 5.11: Excel® link pasted in Word®

User prompts and messages

Adding user prompts and messages helps to make your spreadsheets more user-friendly and also to reduce errors, as the prompts help inform users what they should be doing to the various parts of your data model. As already seen in the Ease of use techniques section, prompts can be added to data validation.

Macro code can include message boxes which pop up when that part of the code is run if you think it useful to bring something to the user's attention at that point, for example to make sure that a data file is where it is expected to be and ready to import into the data model.

Layout and presentation

Excel® offers many options to help you choose the layout and presentation of your data model. You can:

- change the font size and style
- merge cells
- select colours
- include borders
- use shading
- use conditional formatting.

All these options require you to highlight or select cells before formatting them. There are other layout and presentation options such as graphics, headers and footers which require other approaches (see Figure 5.12).

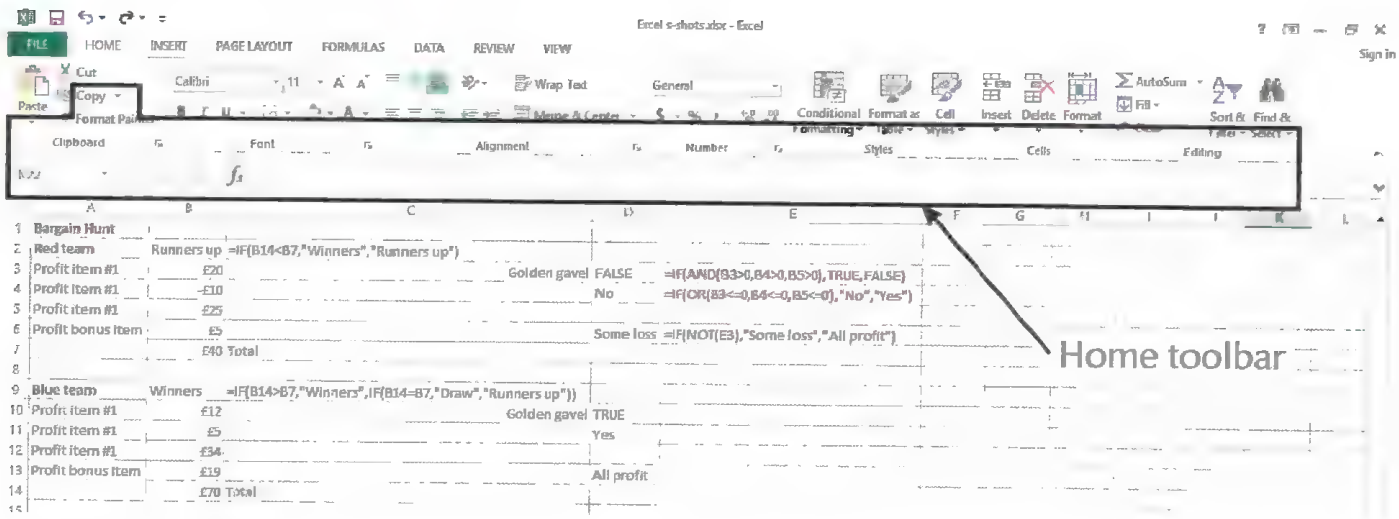


Figure 5.12: Home toolbar formatting

Font size and style

Having the correct font size and style is essential for emphasising the more important text in a spreadsheet and for making the data model more easily understandable. Using a larger font size makes the text more prominent. Using different styles can be used to differentiate between fonts of the same size: **bold** is more prominent than ***bold italic***, which is more prominent than 'normal', which is more prominent than *italic*. **Bold** and ***bold italic*** are often essential when light coloured text is shown over a dark background because, otherwise, this can be very difficult to read.

Excel® gives some choices for copying a font size and style to other cells with the format painter or with paste special. To use the format painter, start by copying the cell with the formatting that you want to use, click on the Format Painter button and then click

on the cell that needs the new format. You can do this to many cells by double clicking on the Format Painter button; this will remain active until you tap the <Esc> key.

To use paste special, you also start by copying the cell with the formatting that you want to use. You can now highlight the cell(s) you want to format, then bring up the dialog box by either clicking on the Paste Special button (in the right click menu) or typing <Alt> <E> then <S> (see Figure 5.13). Select Formats then OK. The paste special can be repeated to format other cells without needing to re-copy the original cell. When completed, you tap the <Esc> key to clear the copy.

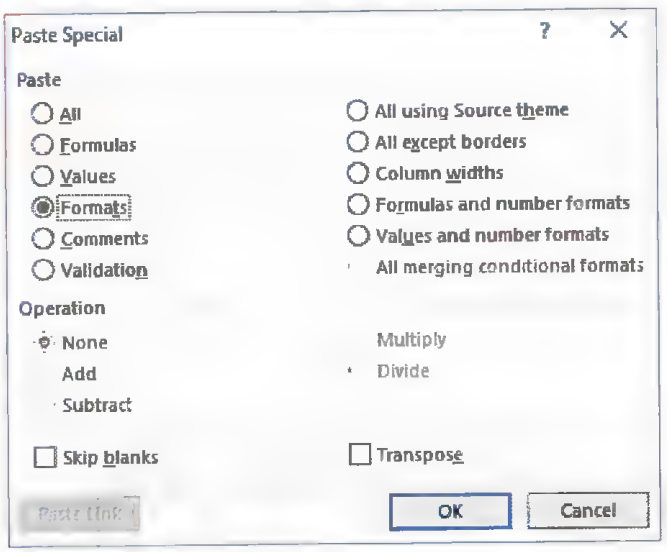


Figure 5.13: Paste Special dialog box

Cell layouts

Colour

Colours are easily set by selecting the cells to be formatted and then clicking on the Font Colour or Background Colour drop-down menus to choose the colour.

Merging cells

Merging cells can be useful for centring a heading over several cells. Select the cells to be merged and then click on the Merge and Centre drop-down menu in the Home toolbar. You will now have three merge choices plus the option to unmerge.

- Merge and Centre will merge the highlighted cells, and any text in the leftmost cell will be centred in the merge.
- Merge Across will merge the highlighted cells and any text in the leftmost cell will keep its alignment in the merge: for example, if it was right aligned before, then it will end up right aligned in the merged cells.
- Merge cells is similar to Merge Across in that the alignment is kept. It differs if more than one row of cells are selected for the merge in that they are merged into one cell, whereas Merge Across merges each row.

For all of these merge options, text in the (top) leftmost cell is kept and text in the other selected cells is lost.

Borders and shading

Borders and shading can be set after selecting cells by bringing up the dialog box (see Figure 5.14). Either right click on them to select Format or click on the Borders and Shading drop-down menu in the Home toolbar to see the Borders menu, where you can click on More borders. To use the Border tab of the Format Cells dialog box, select the style and colour you want and then click on one of the Presets, click in the Border diagram or use a Border button.

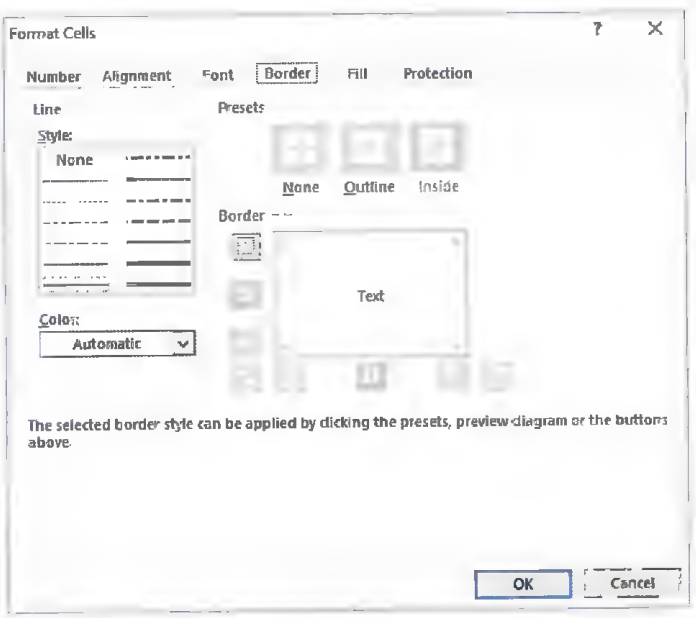


Figure 5.14: Border dialog box

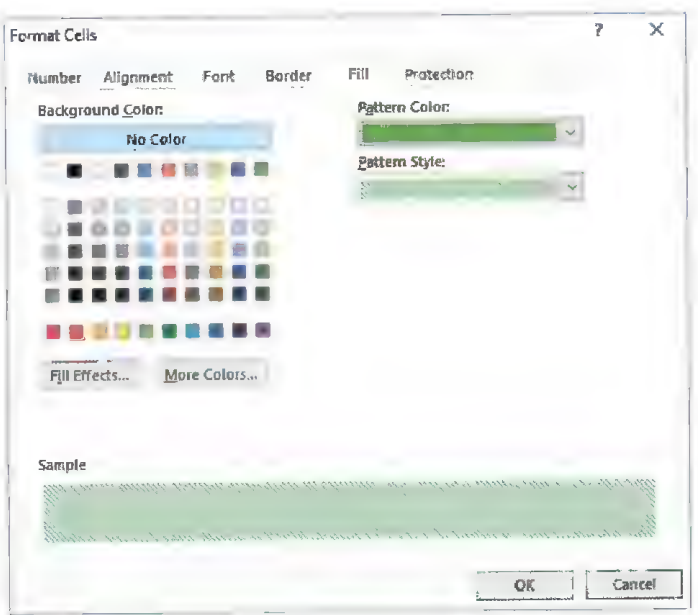


Figure 5.15: Fill dialog box

See Figure 5.15 for the background and shading options. This is shown by clicking on the fill tab of the Format Cells dialog box. Here you can click on a background colour or choose a pattern fill colour and style. The OK button completes any choices using this dialog box.

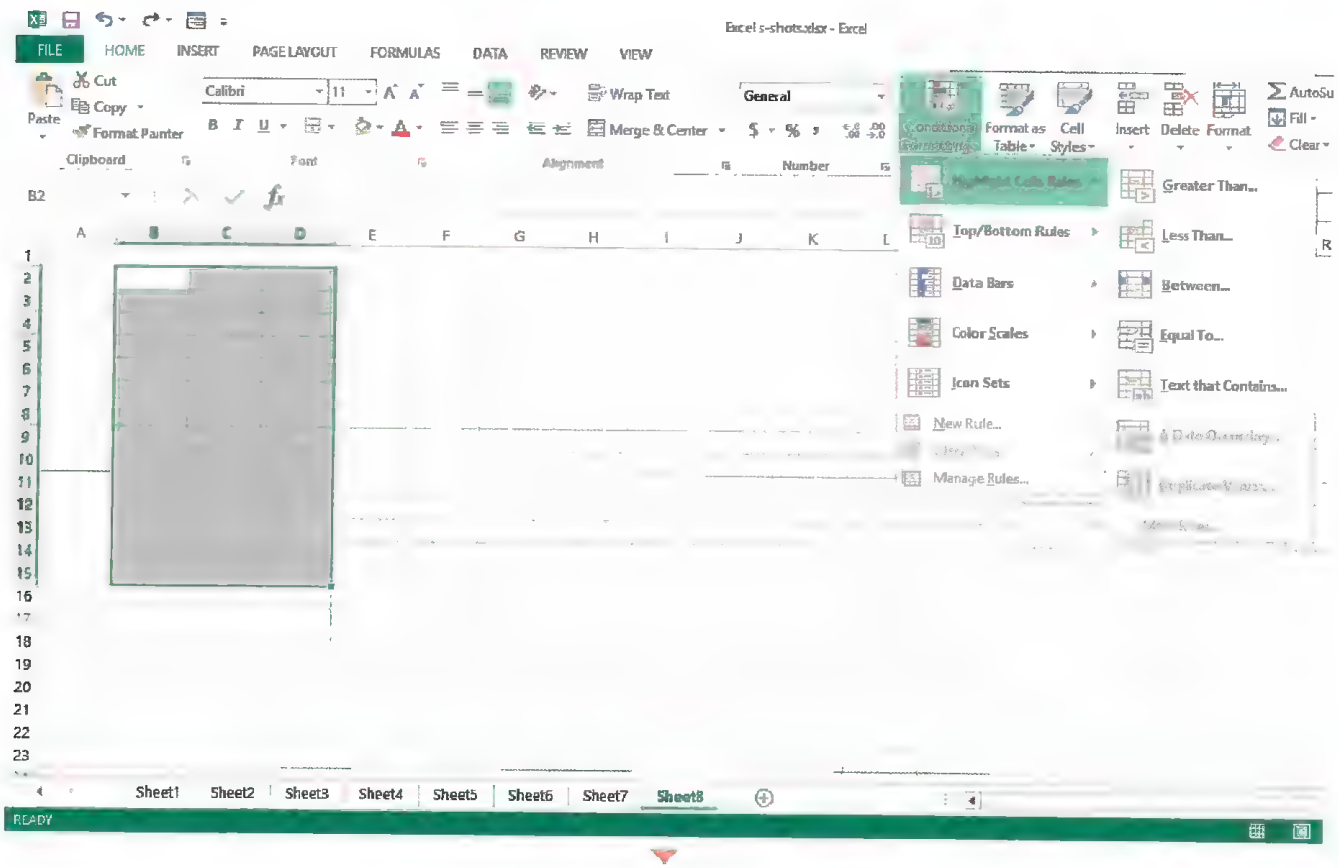
Conditional formatting

When you have a large data set and need to be able to identify certain items of data, for example especially large or small values, conditional formatting is very effective. Select the cells and then use the Conditional Formatting button in the Home toolbar to bring up the options.

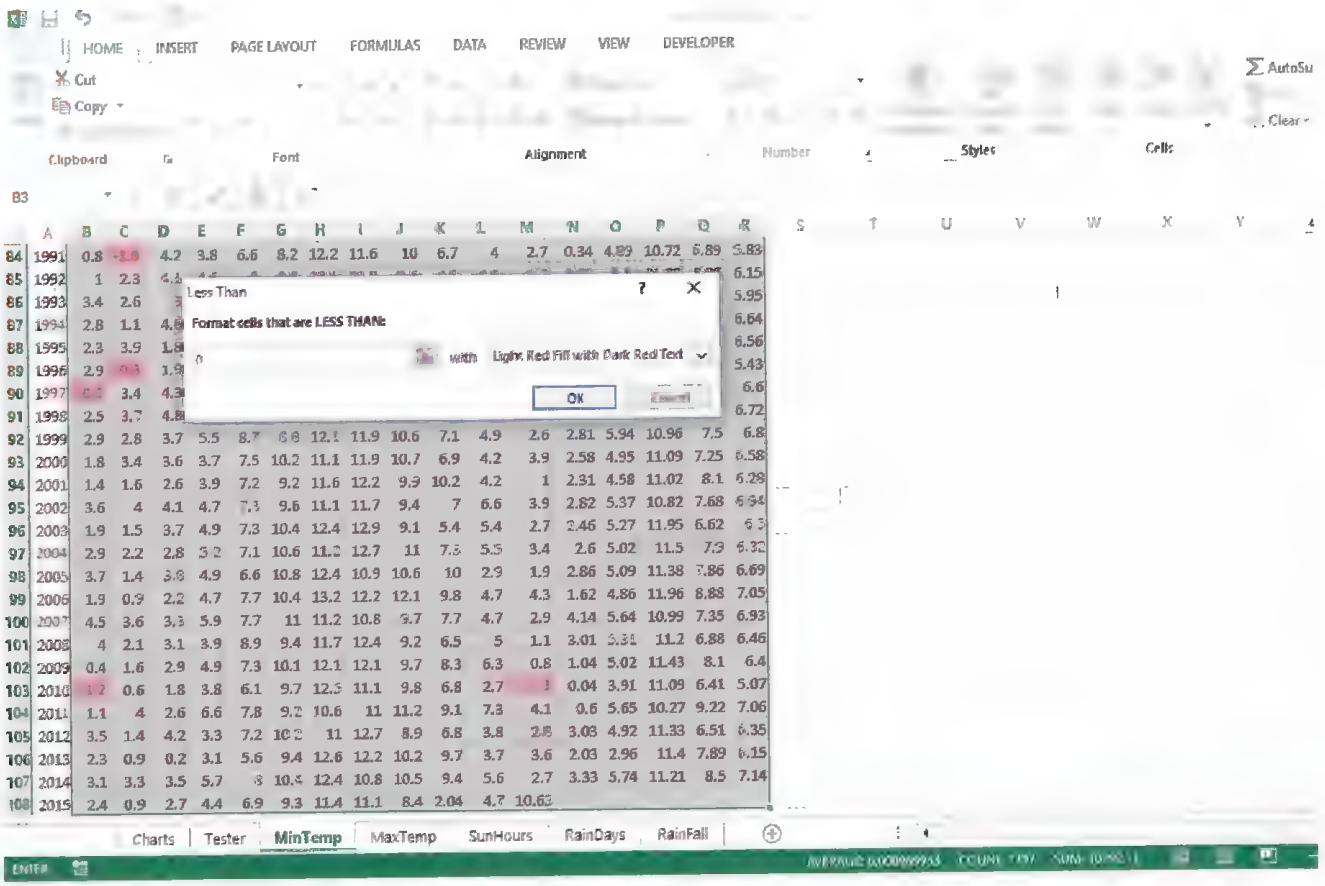
There are many conditions you can apply to format cells that match your criteria (see step 1 in Step by step: Setting up conditional formatting). You can have several conditional formatting rules on the same range of cells, so that you can look for different things such as below a chosen value or above average. This can result in almost any formats for matching cells including font, bold, italic, colour and background. Many spreadsheets use red for bad data and green for good, but anything is possible as long as it matches the user's requirements.

Step by step: Setting up conditional formatting

1 This shows how to enter conditional formatting to colour cells red if they are below a given value. Select the cells where you want the conditional formatting and then use the Conditional Formatting drop-down menu in the Home toolbar to show the available options.



Select the Highlight Cells Rules then Less Than. Excel® shows the Less Than dialog box where you type '0', the number that triggers the conditional format. The default for less than is red. Other colours for cells that meet the trigger can be chosen from the drop-down menu. The OK button completes setting up this conditional formatting.



Headers and footers

It is often useful to add headers and footers, especially when a spreadsheet will cover several pages if printed. Even a single page print will benefit from a footer with fields giving information such as filename or date and time of printing.

Figure 5.16 shows the dialog boxes involved in setting a custom footer. To start, use the Sheet Options drop-down menu in the Page Layout toolbar to bring up the Page Setup dialog box. From here, select the Header/Footer tab then click on Custom Footer to bring up the Footer dialog box.

The Footer dialog box has three sections for the left, centre and right of the footer. You can click inside any of these sections then mix typing and fields from the buttons (Figure 5.16). For example, the left section has a field for Page (number) then 'of' was typed then the field Pages (number of pages). A header can be set using these dialog boxes in just the same way. The OK buttons complete these dialog boxes to save your changes.

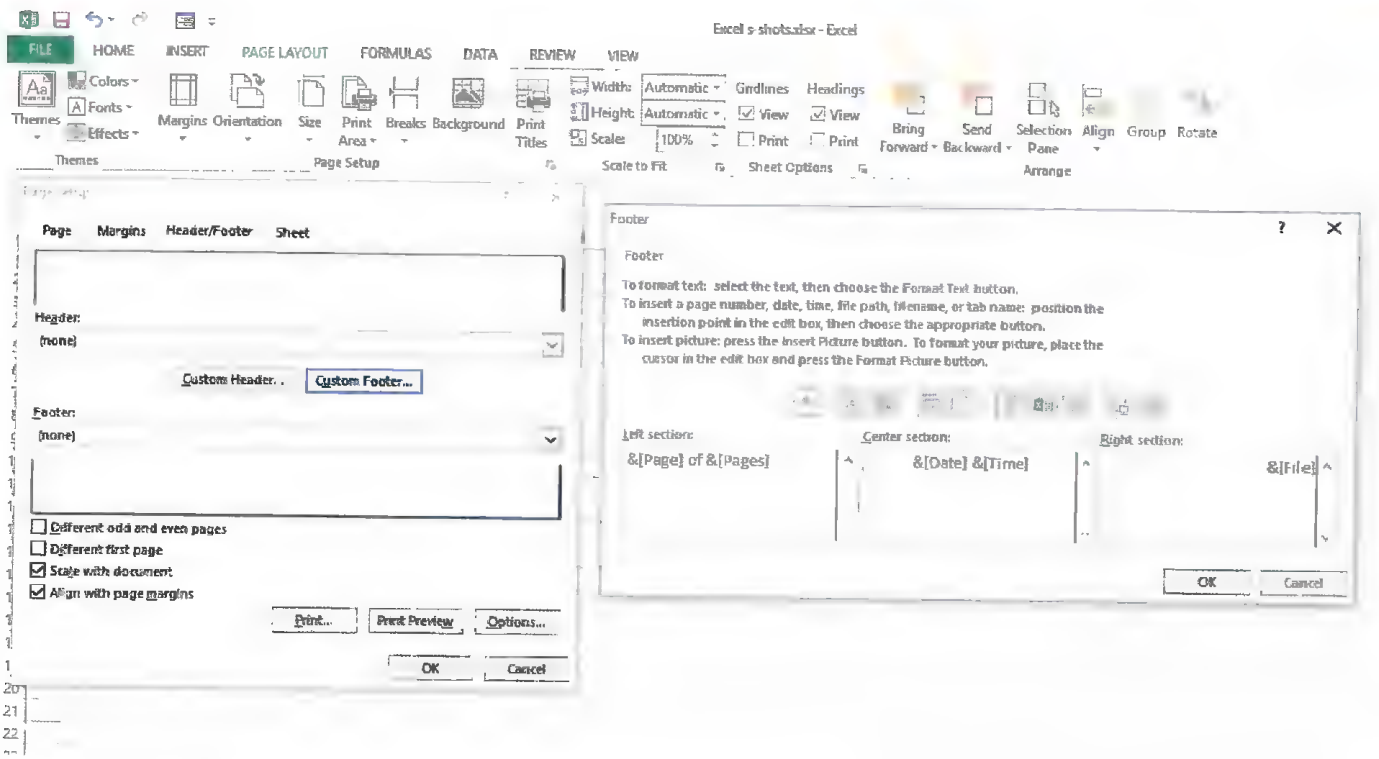


Figure 5.16: Excel® footer

Graphics

You can use graphics other than charts to add diagrams, photos, logos and so on to your spreadsheet. The simplest method for this is to copy the graphic and then paste into the spreadsheet. Once pasted, you can use the mouse to move and resize the graphic. Right-clicking on a graphic object allows you to format any shadows, edges or other aspects of the graphic, for example a line around it.

You can also find a large collection of graphical objects under Shapes in the Insert toolbar, which allows you to enter rectangles, lines or call-outs to your spreadsheet.

Output requirements

Your data model needs to meet the client's requirements in terms of worksheet layout and output from charts and graphs. Your design documentation defines what was agreed with the client for each of these features.

Worksheet layouts

The worksheet layout design should include some detail about what it will be used for and a plan for how the data and calculations will be grouped and kept. Graphics can add meaning to the data model and logos will personalise it. Colours, borders and shading help to make the information more accessible and can also match any corporate style that the client has. You should refer to your designs for all of these aspects of the worksheet layout while implementing your data model to ensure that it meets the client's needs.

Charts and graphs

The types of charts and graphs that you will use for your data model should have been outlined in the planning documentation. You might need to tweak these designs when your actual data is used to produce the charts, to help you communicate your findings more clearly.

Testing the data model solution

You will use both formative and summative testing to test your data models. You should include tests for functionality and user acceptance.

Formative testing is carried out as the model is developed. Each calculation and function should be checked after it is entered to confirm that it gives the expected result.

Summative testing is carried out when the model is completed to confirm that it works as expected.

Testing

Your data model needs thorough testing to establish whether:

- the solution meets all the requirements of the functional specification
- the **underlying logic** is correct
- all the functions and formulae work correctly.

All this testing should be recorded in the test log with outcomes and corrective actions should be taken.

Key term

Underlying logic – how the spreadsheet will process the data into information. This is more of an overview of what the processing is to achieve, rather than the detailed calculations which actually implement the underlying logic.

Meeting the requirements

You will need to check that your data model solution meets all the requirements of the functional specification. To do this, return to the original user requirements and carefully plan tests to ensure that everything that was wanted is present and works well.

A common problem with large projects is that they can sometimes lose sight of what was actually required at the start of the work. For example, a data model might be required to predict monthly sales, but implemented to predict weekly sales.

Underlying logic

Testing is needed to check that the underlying logic of the model is correct. You will need test data with the expected outcomes pre-calculated so that, when you test the model, you can compare the actual results with what was expected.

The underlying logic should be tested with data that would be expected to produce a wide range of outcomes so that you can confirm that the model works as expected.

Functions and formulae

All the functions and formulae in the spreadsheet must work correctly, so (as far as practical) each must be tested. If there are a large number of calculations which have been copied from a row or column, then you can carefully check the originals and the first few copies and the last few for correct working.

Another way to check a large number of calculations is to use the Show Formulas button in the Formulas toolbar, which enables you can see the calculations rather than their results. This test is useful because you can see if anything looks wrong, for example a number in amongst the formulae (where someone has typed over a calculation).

Other factors

During testing, there are other factors to consider. These include:

- the selection and use of appropriate test data
- finding suitable users for solution testing
- gathering feedback from users
- completing the test documentation.

Test data

As already explained, the selection and use of appropriate test data makes a big differenc e to how well your data model is tested and therefore to how useful it is.

Valid data is needed to prove that calculations produce correct and reliable results. Valid data consists of any test data which is expected to produce a good output.

Invalid data is used to find out what this does to the data model. Invalid data is test data that is obviously wrong, for example a date entered as 35/5/2016. Ideally, it would be rejected from the checks you have set on the data entries.

Erroneous is similar to invalid, except that it should pass as valid, even though it is wrong, for example a misspelt word or a number that was wrongly entered. This should be handled by the model without crashing or producing system error messages.

Extreme data should be used to check the limits of the model. Ideally, this would be at both the low and high limits of what is acceptable, just inside and just outside. As in all the other tests, the model should respond to these extremes as expected. If it does not, then any differences need to be corrected.

Users

Users are an important part of your testing, so you need to select suitable users to test your solution who are able to communicate their findings to you. Gathering feedback from users will enable you to respond to their findings.

PAUSE POINT How will you test the data model for correctness, functionality and acceptance?

Create notes under the subheadings correctness, functionality and acceptance on what you would test.

In addition, state what your test data will need to achieve for each type of test.

You should be particularly interested in how users find the effectiveness, presentation, performance and users of your data model.

Test documentation

Your test logs will document what you did and what happened: that is, which tests were passed and the remedial actions you took to rectify any failed tests.

Try to design this test documentation so that it looks professional and is able to communicate the thoroughness of your testing and the reliability of the model. To do this, the documentation must be completed during every test and the outcome must be carefully recorded.

Link

For more on test documentation see Test plans.

Reviewing and refining the data model solution

As the project nears completion, there will still be some improvements or refinements that you can make to the data model, in order to fully meet the client requirements.

Refining the model

Refining the model may need to take into account any issues raised during testing. Each of these issues must be carefully considered and evaluated. How much impact would they have on improving your spreadsheet? You will have a limit on how much time is available, so you will need to weigh up the expected benefits against the practicalities of implementing them. For example, calculations in the model might be optimised to simplify the inputs, processing or outputs.

There could also be an opportunity to refine the model to take account of the feedback and client requirements. Other people can be a valuable source of ideas and can also identify problems in the model that you did not notice or anticipate. You will again need to evaluate whether it is practical to make refinements based on the feedback at this time.

Document any of the issues raised that were not practical for you to implement at this time as ideas for extending your model.

Reflect

Evaluating the outcomes of your work is necessary to help you to produce high quality, justified recommendations and decisions. Your evaluations are essential to achieve the highest grades for this unit and to maximise the quality of your data modelling spreadsheet solution.

Evaluate what you could have done better and what improvements you could make if you had time.

Extending the model

Always include in your documentation any ideas you have which could be used to extend the data model. This section will be useful to communicate an evaluation of your understanding of the user requirements, the current value of your spreadsheet solution and the future potential of your data model.

The future potential of your data model could be realised by extending the model. What could be added to the model to make it more useful? Would more (or less) user inputs offer more possible outcomes or make it easier to use? Could the meanings of the data model outputs be enhanced using charts or different formatting?

PAUSE POINT How can a data model be optimised?

Create notes under the subheadings testing, user feedback, inputs, calculations and outputs.

Identify an appropriate use for each of the different Excel® chart types.

Reflect

Your communication skills using different media will have improved during this unit. The ability to convey your intended meanings using written and verbal communications will have developed in many ways as you planned and developed your project.

Written communications have included emails, design documentation, recording documentation, reports and the visual aids you used for presentations.

Verbal communications have been practised in both informal and formal one-to-one and group situations. You will have improved your use of tone and language for verbal and written communications, to convey intended meaning and make a positive and constructive impact on your audiences.

Your verbal communications should now use a positive and engaging tone with appropriate choices of technical and vocational language which is suitable for your intended audience and avoids use of jargon, where possible.

There have been many opportunities to respond constructively to the contributions of others. You will have been supportive of their views and opinions, managing their contributions so that they all had opportunities to contribute.

You will have responded positively to any objections, listened carefully and understood the points raised. The expectations of stakeholders will have been managed by your planning and responses to user comments. Any conflicts will have been resolved by understanding the issues and negotiating solutions.

Consider how you have personally developed your written and verbal communication skills during this unit.

Reflect

You will need to evaluate the targets that you set for yourself against how well you achieved them in order to obtain insights into your own performance. Evidence of this will be needed for the higher grades. This skill is exercised by many professional high-achievers to enhance their workplace performance.

Assessment practice 5.2

B.P3 B.P4 C.P5 C.P6 C.P7 B.M2 C.M3 BC.D2 BC.D3

Your job as junior programmer at W-Ready is going well and the purchasing manager, Ellie, has accepted your report evaluating the role of spreadsheet data modelling in decision making.

Ellie would like the data model to include collections of real data on weather, such as minimum/maximum temperatures, hours of sunshine, average rainfall and number of rain-free days, along with predictions for the future, so as to produce useful information on which she can base her purchasing decisions.

She particularly wants to be able to compare the temperature, sunshine and rainfall for any month she selects and to control how many years of back data are used to make predictions for the next two years.

Plan

- Do I need to be working with anyone else? If so, what is my role? What am I contributing?
- Do I have any existing knowledge around the task at hand?
- What resources do I need to complete the task? How can I get access to them?
- How much time do I have to complete the task? How am I going to successfully plan my time and keep track of my progress?
- What aspects of the task do I think will take the most/least time? How will I balance these?

The programming team are to produce a spreadsheet for Ellie, and your team leader, Salim, has given you the task of producing the following documentation for the project.

- A functional specification to clarify and define the data model.
- Design documentation including:
  - designs for the data model meeting client requirements
  - review of your designs including feedback from Ellie and others in the programming team
  - improvements to your designs
  - justification of the decisions you made
  - how your chosen design will fulfil its purpose and client requirements.
- Testing logs to record your spreadsheet tests and any actions you took to resolve problems.

You will need to use appropriate methods to inform Ellie and Salim of the design improvements to confirm that they both agree with these changes to the specification. These will be documented as screenshots, recordings or witness statements.

You are then asked by Salim to create and develop a data model to fulfil your design using a range of appropriate and advanced spreadsheet features and functions.

Carry out comprehensive testing and seek user feedback to refine and improve your data model. Make sure that the model is thoroughly tested to prove that it meets Ellie's requirements and keep your testing log up to date.

Produce a report on your spreadsheet to:

- evaluate the effectiveness of the alternatives you considered along with ways they could be improved if you repeated the task
- review the extent to which the data model meets client requirements
- explain how you optimised the data model to meet client requirements.

This report will evaluate the decisions and processes that you applied throughout the design, development and testing stages along with their impact on the effectiveness of the final solution. You should evaluate your alternatives along with their impacts and consequences and justify the selection that produced the best decision or compromise.

Your evaluation should be a systematic and accurate review of your own skills and performance and how this improved the effectiveness of the solutions. Identify where you took responsibility for your own work (for example, identifying potential issues and how you resolved them, the improvements you made and how you kept the work safe and secure) and explain how you made responsible use of any copyrighted content.

Document any creative or innovative approaches you took to this problem solving. Evaluate your final design and produce well-considered, justifiable suggestions for future improvements to your data model.

Do

- I can seek others' opinions.
- I am open to change.
- I can make connections between what I am reading/researching and the task, and identify the important information.
- I am recording my own observations and thoughts.
- I am recording any problems that I am experiencing and looking for ways/solutions to clarify queries.
- Am I utilising all the support available to me?
- What am I struggling with? Do I know how to overcome this?
- I can question my own learning approach.
- I can identify when I have gone wrong and adjust my thinking/approach to get myself back on course.

Review

- I can explain which elements I found easiest and hardest.
- I can explain how I would approach the hard elements differently next time.
- I can describe my thought processes.
- I can explain the skills that I employed and the new ones I have developed.
- I can make informed choices based on reflection.
- I can explain what I have learned and why it is important.
- I can explain what success looks like.
- I can use this experience in future tasks/learning experiences to improve my planning/approach and to monitor my own progress.
- I can identify how this learning experience relates to future experiences (ie in the workplace).
- I realise where I still have learning/knowledge gaps and I know how to resolve them.

Further reading and resources

Alexander, M. (2015). *Excel Macros For Dummies*. New Jersey: John Wiley & Sons.

Harvey, G. (2013). *Excel 2016 For Dummies (Excel for Dummies)*. New Jersey: John Wiley & Sons.

Laing, R., Hawkins, R. and Mayne, M. (2015). *Microsoft Excel Basics*. London: Flame Tree Publishing.

Websites

Microsoft® Office® help and training: **www.support.office.com**
This is a comprehensive website supporting Excel® and other MS apps.

Stack Overflow: **www.stackoverflow.com**
This website provides Excel VBA macro programming support.

Excel Easy: **www.excel-easy.com**
This website offers general Excel support.

Excel Functions: **www.excelfunctions.net**
This has a good choice of Excel step-by-step support.

THINK FUTURE

Billy Stephens

Systems analyst in a large bank.

I've been working for a large bank for ten years now, progressing from junior programmer to systems analyst, responsible for a team of six programmers. My role is varied, interesting and busy. I have to communicate with users, their managers and my programming team to understand the users' needs so as to design suitable programs that my team will then produce.

Data modelling is part of my role. We produce spreadsheets that meet the users' needs. This is usually a much smaller task than a full programming project, which would require a language such as C#, and it can be delivered a lot more quickly. There is a steady demand for small spreadsheet-based data models so that users can explore 'what if' scenarios with a lot more flexibility than our program projects usually deliver.

Our spreadsheet models often use data that is downloaded from our main systems so that users have up-to-date information to examine. Obviously, we try to avoid our users needing to type in large amounts of data as this is time consuming, error-prone and unnecessary. We usually set up some cells so they can explore percentage changes in the data and include charts to help them easily understand the information that our models produce.

Focusing your skills

Data analysis

Data and information are everything in the modern workplace. Computer systems are used by every organisation to input, store and process data and output information. It is important to be able to recognise the data required to produce useful information along with the processing required to do this. There are very good opportunities for people who are able to identify such data to pursue careers as systems analysts.

Here are some simple tips to consider.

- What data is available?
- Can the data be used as it is? That is, does it need to be typed in, or can it be imported into your data modelling app?
- Does the data need more content? That is, by typing or by combining with other data.
- Make sure that you understand how the processing is going work.
- Make sure that you know what information is required: that is, the content and how it will be presented.

Spreadsheet creation

The spreadsheet is a great app for creating data models. There is flexibility on how the model can be structured because of the enormous collection of functions available and there is a wide range of possible ways of showing the information using charts and cell formatting.

Spreadsheet creation can be a good route towards many career opportunities involving systems analysis and programming.

Here are some useful tips.

- Practise using the spreadsheet to become familiar with the calculations, functions, charts and formatting options before planning your data model so that you know the options open to you.
- Plan your spreadsheet before you start work on the data model.
- Produce a test plan to clarify your understanding of what is required from your data model.
- Check every calculation as you enter it.
- Use paste special to control how you copy and paste, especially values, formulae or formats.
- Save your work regularly and use save as to make a copy of your model before making major changes.
- Check that printing works well, is easily readable and does not spill over onto unwanted pages.

Getting Ready for Assessment

Kerri is working towards a BTEC Level 3 National Extended Certificate in Information Technology. She was given an assignment with the following title: 'Investigate data modelling and decision making' for learning aim A. She had to write a report with these sections:

- stages involved in the decision-making process for data modelling with explanations
- features of spreadsheet software that can be used to support the decision-making process and how they can be used
- an analysis of how spreadsheet software features help decision making
- an evaluation of how spreadsheet software features contribute to the decision-making process.

Kerri shares her experience below.

How I got started

I set about using several search engines to research spreadsheet data modelling on the internet. The websites I thought most useful were bookmarked into the two browser folders I made for data modelling and for spreadsheets.

Next, I visited the central library to check out the books available for some background reading. This was quite difficult as most of the titles were about databases, but I did find one which I took home for a fortnight. I quite like books because I find flipping through the pages a quick and effective way of skipping past what I don't want and finding what I need.

I met up with one of the IT team at my centre for half an hour to talk through what they do with spreadsheets to support their managers. This introduced me to something called the Cube which is a pivot table they use to hook into their network data sources.

How I brought it all together

I created a Word® document for my report with headings for each of my sections. I also added an introduction, a summary and a bibliography section. I typed my name into the header and page numbers into the footer.

The stages involved in the decision-making process for data modelling section just involved reading my research and typing it up. My section on the spreadsheet features needed quite a lot of hands-on time with Excel® to get familiar with what could actually be done with this app. I focused on four features, which I illustrated with my own screenshots.

- The analysis of spreadsheet software features helping decision making section was a mixture of my hands-on experience with Excel® and the discussions around using the Cube.
- My evaluation of how spreadsheet software features contribute to the decision-making process simply recycled my previous research with extra depth and my thoughts on how these work together.

What I learned from the experience

The biggest surprise was meeting the IT team member and seeing what can actually be done in a modern system, especially the way charts were used to present information. My report writing is a little stronger now and I'm a lot better at using Excel®.

Think about it

- Do you have a plan with timings so that you can complete your assignment by the submission date?
- Is there a workplace or organisation that you can visit as part of your research?
- Which features of a spreadsheet are most suitable for supporting a data model?